Network meta-analysis with integrated nested Laplace approximations

---

Master Thesis in Biostatistics (STA495)

by

Burak Kursad Gunhan
13-730-023

supervised by

Prof. Dr. Leonhard Held, University of Zurich
Rafael Sauter, University of Zurich

Zurich, February 2016

Contents

Abstract iii

1 Meta-analysis 1
- 1.1 Introduction . . . 1
- 1.2 Meta-analysis: The analysis of analyses . . . 2
- 1.3 Statistics of meta-analysis . . . 2
- 1.4 Meta-regression . . . 15

2 Network meta-analysis 21
- 2.1 Introduction . . . 21
- 2.2 Some terminology in network meta-analysis . . . 23
- 2.3 Statistical models for network meta-analysis . . . 26
- 2.4 An application (Hip dataset) . . . 31

3 Design-by-treatment interaction model 41
- 3.1 Introduction . . . 41
- 3.2 Cycle inconsistency and design inconsistency . . . 42
- 3.3 The Design-by-treatment interaction model . . . 45
- 3.4 An application (Smoking dataset) . . . 46
- 3.5 The Lu-Ades vs the Jackson model . . . 50
- 3.6 An application (Hip dataset) . . . 53

4 Conclusions and future research 57

A R-code and BUGS/JAGS-code 61
- A.1 Pairwise meta-analysis models . . . 61
- A.2 The consistency model . . . 68
- A.3 The Lu-Ades model . . . 71
- A.4 The Jackson model . . . 73

Acknowledgments 77

References **79**

Network meta-analysis with integrated nested Laplace approximations

Burak Kursad Gunhan [1]

Abstract

This thesis investigates how to perform inference with different approaches in meta-analysis models as well as in regression-type meta-analysis models named meta-regression. Chapter 1 contains an introduction to meta-analysis as well as different statistical models and estimation techniques for meta-analysis. Also, a recent Bayesian inference method named integrated nested Laplace approximations (INLA) is used for making estimations in meta-analysis. Chapter 2 contains a motivation for a broader type of meta-analysis called network meta-analysis (NMA). This chapter introduces two models, namely the Lumley and the Lu-Ades models, for NMA and shows how INLA apply to those models. Chapter 3 starts with the discussion of the distinction for different types of inconsistency in the network, namely the cycle inconsistency and the design inconsistency. Then, the design-by-interaction model using random inconsistency parameters, the Jackson model, is introduced. This chapter continues with showing how INLA can be used as an inference method for the Jackson model. Also, Chapter 3 shows that the Lu-Ades models depend on the treatment ordering while the Jackson model do not for an application. All analysis was performed in the R programming language (R Core Team, 2015). Three different applications were used to demonstrate the use of INLA and other methods. Appendix includes the R-code which are used to obtain the results in the main text and the BUGS/JAGS-code to fit the consistency and the Jackson model with MCMC. Also, an R function, **meta.inla**, which is developed to implement the models introduced in Chapter 1 with INLA is given.

Key words: Meta-analysis; integrated nested Laplace approximations; network meta-analysis; design inconsistency, design-by-treatment interaction model

[1] *burakgunhnan@gmail.com*

Chapter 1

Meta-analysis

1.1 Introduction

> ...The work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out (Rayleigh, 1885).

So said the professor of physics at Cambridge University in his presidential address to the 54th meeting of the British Association for the Advancement of Science held in Montreal in 1884.

The importance of the process of review of evidences from different studies is widely recognized. However, one of the most crucial issue is the characteristics of such review or synthesis process. According to Hothorn and Everitt (2014) "the classical narrative review" of different studies can be very misleading, because of the possible biased selection of evidence and "the emphasis placed upon it by the reviewer to support his or her personal opinion". On the other hand, an alternative and systematic way of such synthesis that has become famous in recent years is *systematic review*, defined in *A Dictionary of Epidemiology* (Last *et al.*, 2001) as follows:

> The application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic.

In this thesis, my purpose is merely concentrating on *meta-analysis* which is the quantitative part of a systematic review, although there exists qualitative components of a systematic review. In this chapter, I will describe first meta-analysis in general, then different techniques within the statistics of meta-analysis and finally regression-type methods in the meta-analysis context will be introduced.

1.2 Meta-analysis: The analysis of analyses

Gene Glass coined the term meta-analysis in 1976 in a presidential address stressing the need for better synthesis of research results (Chalmers *et al.*, 2002), and it is defined in the *Cambridge Dictionary of Statistics in the Medical Sciences* (Everitt, 2002) as follows:

> A collection of techniques whereby the results of two or more independent studies are statistically combined to yield an overall answer to a question of interest. The rationale behind this approach is to provide a test with more power than is provided by the separate studies themselves. The procedure has become increasingly popular in the last decade or so, but is not without critics, particularly because of the difficulties of knowing which studies should be included and to which population final results actually apply.

Meta-analysis gives the systematic review an objectivity by the help of its quantitative nature which embodies different statistical techniques. However, before using statistical methods for combining the results of different studies, maybe the first and foremost question which should be asked by any reviewer is that "Where do studies come from?". Moreover, as pointed out in the above definition, main criticisms of meta-analysis methods are still about this question of *study selection*.

Study selection should be regarded as one of the most fundamental aspect of a meta-analysis. The selected studies should be "sufficiently homogeneous" regarding in and exclusion criteria. However, the thesis is aimed to focus on the statistics of a meta-analysis rather than study selection, see Garg *et al.* (2008), Petitti (2000) and Harvard and Lau (1993) for discussions about concepts of selection criterion and the importance of this aspect. Also it is worth noting The Cochrane Collaboration which is a specialized initiative for the systematic review processes including the study selection (details of the structured process of Cochrane systematic reviews are available through their website: http://www.cochrane.org).

1.3 Statistics of meta-analysis

Before explaining the statistics of meta-analysis, I introduce an illustrative dataset which will be used through all techniques as it is easier to show concepts and inference methods with an example and give definitions of some terminology from the field of biomedical research.

1.3.1 TB dataset

Bacille Calmette Guerin (BCG) is one of the most widely used vaccination against tuberculosis (TB) in the world. And the data reported by Coldlitz *et al.* (1994) consists of

13 clinical trials of BCG vaccine each investigating its efficacy in the prevention of TB. The TB dataset which includes the number of subjects suffering from TB with BCG vaccination or without BCG vaccination is shown for each study in Table 1.1. Furthermore, values of two other variables for each trial, namely, the geographic latitude of the place where the trial was undertaken and the year of publication are given.

Table 1.1: TB dataset. The number of TB cases after vaccination with BCG (TRTTB), the total number of people who received BCG (TRT), the number of TB cases without vaccination (CONTB), the total number of people in the trial without vaccination (CON) as well as the geographic latitude of the place where the trial was undertaken (Latitude) and the year of publication (Year)

Trial	TRTTB	TRT	CONTB	CON	Latitude	Year
1	4	123	11	139	44	1948
2	6	306	29	303	55	1949
3	3	231	11	220	42	1960
4	62	13598	248	12867	52	1977
5	33	5069	47	5808	13	1973
6	180	1541	372	1451	44	1953
7	8	2545	10	629	19	1973
8	505	88391	499	88391	13	1980
9	29	7499	45	7277	27	1968
10	17	1716	65	1665	42	1961
11	186	50634	141	27338	18	1974
12	5	2498	3	2341	33	1969
13	27	16913	29	17854	33	1976

In the biomedical research framework, *arm* is used for a group of people who receive the same treatment. *Treatment arm* is the arm which receives a certain treatment, while *control arm* is the arm which receives the standard treatment or placebo. Therefore, in our example the second and third columns in Table 1.1 belong to the treatment arm whereas the fourth and fifth columns belong to the control arm. Furthermore, *treatment effect* is simply a way of quantifying the size of difference between two groups, here between treatment arm and control arm. There are different measures of treatment effect such as odds ratio, log odds ratio, relative risk or difference in means.

When we consider the TB dataset, our main purpose in collecting the results from the controlled trials of using BCG to prevent TB was to estimate the overall log odds ratio, θ, which is, more precisely,

$$\theta = \log\left(\frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}}\right) \tag{1.1}$$

where π_1 is the ratio of suffering from TB in the treatment arm and π_0 is the ratio of suffering from TB in the control arm. At this point, we will introduce first a simplistic approach that is simple pooling, then two frequently used methods which are fixed effect and random effects models to achieve this goal.

1.3.2 Simple pooling

Intuitively, as a first method for estimating the overall treatment effect, we can assume the unknown disease risk for treatment arm and control arm do not vary across the different studies, so we can simply pool the whole data and perform the analysis which is called *simple pooling* (Bravata and Olkin, 2001). For the TB dataset, if we simply pool the results, we obtain the results shown in Table 1.2.

Table 1.2: Contingency table of pooled version of the TB dataset. The number of people suffering from TB for each arm (TB) and the number of people do not have TB for each arm (non-TB).

	TB	**non-TB**
TRT	1065	189999
CON	1510	164773

Then, we can obtain the Maximum likelihood estimate (MLE) $\hat{\theta}_{SP} = log[(1065 \cdot 164773)/(189999 \cdot 1510)] = -0.49$ of the log odds ratio and standard error as $\text{se}(\hat{\theta}_{SP}) = \sqrt{(1/1065) + (1/164773) + (1/189999) + (1/1510)} = 0.04$. Therefore 95 % Wald confidence interval for $\hat{\theta}_{SP}$ has limits -0.57 to -0.41. The estimated log odds ratio with 95 % Wald confidence interval can be seen in Figure 1.1.

On the other hand, this procedure is unrealistic and can yield poor results, since it ignores the characteristics of different studies being pooled and hence the analysis is performed as if the data were derived from a single sample. That's why in our following technique, data from different trials are weighted first, then combined, thereby avoiding some of the problems of simple pooling.

1.3.3 Fixed effect model

Instead of the above simplistic procedure, in a *fixed effect model* (Whitehead and Whitehead, 1991), we compute study-specific treatment effects, $\hat{\theta}_i$, but assume that the underlying true treatment effect, θ, does not vary across studies. The model has the form as follows:

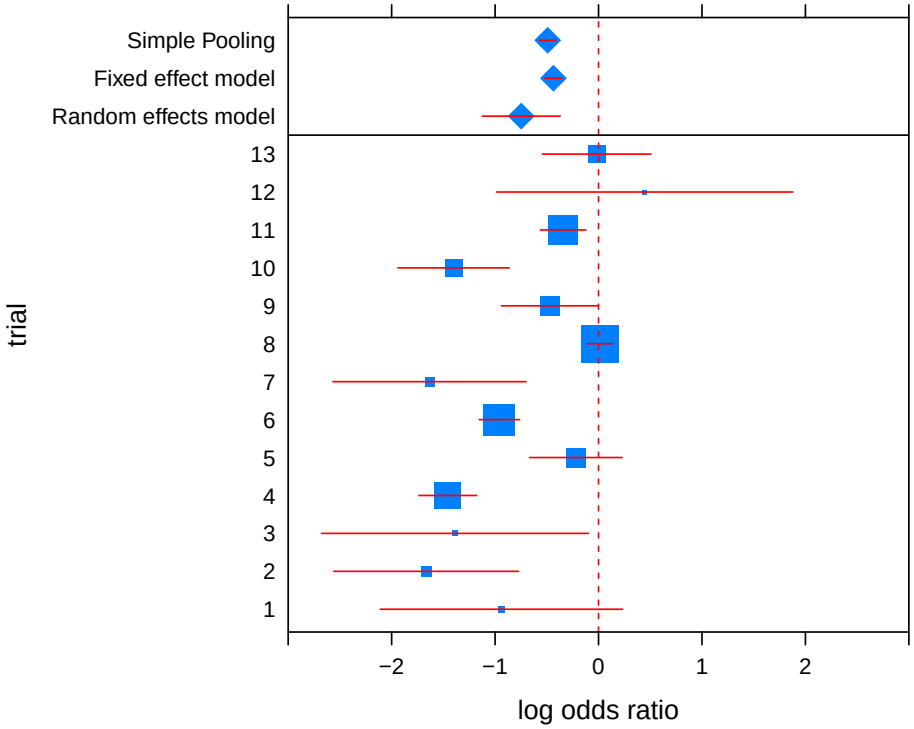

Figure 1.1: 95 % Wald confidence intervals for the trial-specific log odds ratios $\hat{\theta}_i$. The area of the boxes represent the weights given to the trial in the fixed effect model. The diamond figure at the top represent overall log odds ratio calculated by simple pooling $\hat{\theta}_{SP}$ and below represent treatment effect estimate by the fixed effect model $\hat{\theta}_{FE}$ with their corresponding 95 % Wald confidence intervals. Also shown is a 95 % Wald confidence interval for the mean treatment effect ν that is estimated by method of moment approach ("Random effects model").

$$\hat{\theta}_i \sim \mathcal{N}(\theta, \sigma_i^2) \tag{1.2}$$

where σ_i^2 is the within-study variance.

For the estimation procedure, one common way is that the model uses as weighted average of the study-specific treatment effects and the weights being inversely proportional to the within-study variances. This method is called the *inverse variance-weighted* approach. Precisely,

$$\hat{\theta}_{IVW} = \frac{\sum_{i=1}^{k} \omega_i \hat{\theta}_i}{\sum_{i=1}^{k} \omega_i} \tag{1.3}$$

where k is the number of the studies being analysed, $\hat{\theta}_i$ is the treatment effect estimated in the i^{th} study and $\omega_i = 1/\sigma_i^2$. Also, estimated variance of $\hat{\theta}_{IVW}$ is given by

$$\text{Var}(\hat{\theta}_{IVW}) = \frac{1}{\sum_{i=1}^{k} \omega_i} \tag{1.4}$$

In order to analyze the TB dataset, firstly we compute the study-specific log odds ratio estimates, $\hat{\theta}_i$'s, with corresponding 95% Wald confidence intervals by the same methodology which we used in the simple pooling. Then by following the above estimation procedure, we get an overall log odds ratio $\hat{\theta}_{FE} = -0.44$ and with 95% Wald confidence interval from -0.52 to -0.35, so very similar results as in the simple pooling. Figure 1.1 demonstrates all estimated log odds ratios with corresponding 95% Wald confidence intervals.

By using such a model, it is important to realize that we are ignoring possible heterogeneity between studies (or between-study variation) since the fixed effect model has the assumption of common true treatment effect. However, especially when we consider studies from biomedical and social sciences, the assumption of homogeneity between studies can rarely be hold. This is because these studies are likely to have numerous differences, including populations that are addressed, exposures or interventions under investigation are examined (Higgins *et al.*, 2009). Hence between-study variation, in many cases, can be seen as a crucial source of uncertainty that should not be ignored. In the following section, different methods will be discussed for addressing this aspect.

1.3.4 Random effects model

In order to allow for heterogeneity between studies, unlike the assumption of a common true treatment effect in the fixed effect model, a *random effects model* assumes the treatment effect parameter, θ_i, follows a certain distribution which is assumed to be a normal distribution with mean ν and variance τ^2 (Sutton and Abrams, 2001). The random effects model has the form as follows:

$$\begin{aligned} \hat{\theta}_i|\theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \mathcal{N}(\nu, \tau^2) \end{aligned} \tag{1.5}$$

Because of the definition of the model formula, there are two parameters to estimate, namely the mean treatment effect ν and the heterogeneity variance τ^2, unlike the fixed effect model. Although ν may be parameter of primary interest, estimation of τ^2 is just as important, since this variance explicitly describes the extent of the heterogeneity. One important feature of this model must be emphasized that if $\tau^2 = 0$, then the model is exactly same as the fixed effect model. Therefore, the fixed effect model is contained in the random effects model as a special case.

Essentially, inference methods for random effects model can be divided into two categories: classical and Bayesian. In this section we will first discuss the method of moment

and likelihood approaches in classical inference, then describe empirical Bayes and fully Bayes approaches in Bayesian inference. Moreover, in fully Bayes approach we will show Markov Chain Monte Carlo and integrated nested Laplace approximations methods. We will show the results of the implemented techniques to the TB dataset in Table 1.3 at the end of the section. Also we refer to Higgins *et al.* (2009) for descriptions and comparisons about inference methods of random effects models.

Classical inference

- **Method of moment approach**

  One classical approach to random effects meta-analysis is that the estimate of ν is the same as given in equation (1.3) but in this case the weights are given by $\omega_i = 1/(\sigma_i^2 + \hat{\tau}^2)$ (Whitehead and Whitehead, 1991). An estimator for the heterogeneity variance τ^2 is given by using a moment-based approach (i.e. the method of moment estimator for τ^2 is used) as follows (DerSimonian and Laird, 1986):

$$\hat{\tau}^2 = \begin{cases} 0 & \text{if Q} \leq \text{k-1} \\ \frac{(Q-(k-1))}{\sum \sigma_i^{-2} - \sum \sigma_i^{-4} / \sum \sigma_i^{-2}} & \text{if Q} > \text{k-1} \end{cases} \tag{1.6}$$

  where $Q = \sum_{i=1}^{k} (\hat{\theta}_i - \hat{\theta})^2 \sigma_i^{-2}$ and $\hat{\theta} = \frac{\sum \hat{\theta}_i \sigma_i^{-2}}{\sum \sigma_i^{-2}}$.

  This methodology is implemented in the R package `metafor` (Viechtbauer, 2010) and we rely on this package for estimation. Specifically by applying the **rma.uni** function by using the argument of `method = "DL"`, moment-based approach can be implemented. The corresponding R-code can be seen in Appendix A.1.1. As a result, we get log odds ratio of -0.75 with a 95% confidence interval of $(-1.12, -0.37)$ and estimated heterogeneity variance τ^2 of 0.37. Note that $\hat{\nu}_{MOM} = -0.75$ in this model is smaller than under a fixed effect model ($\hat{\theta}_{IVW} = -0.44$) and also corresponding confidence interval is substantially wider than the one for the fixed effect model. A comparison between the estimated ν and overall log odds ratios estimated under the fixed effect model can be seen in Figure 1.1.

- **Likelihood approach**

  A likelihood approach for random effects meta-analysis is possible by using a linear mixed model. This technique is suggested and used by Lumley (2002) in a more complicated context, however when we adapt it to our case, we obtain a linear mixed model formulation which reflects a hierarchical structure as follows:

$$\begin{aligned} \hat{\theta}_i | \theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \mathcal{N}(\nu + \gamma_i, \sigma_i^2) \end{aligned} \tag{1.7}$$

This model allows for an additional source of uncertainty by introducing variation through random effects γ_i. Trial specific heterogeneity captured by the random effects $\gamma_i \sim \mathcal{N}(0, \tau^2)$. Here, the heterogeneity variance τ^2 is a measure for the degree of heterogeneity or between-study variation. A crucial component of every random effects model is the assessment of heterogeneity. With this likelihood approach, a large heterogeneity variance τ^2 indicates that there is a between-trial variability exceeding the expected sampling variability for the mean treatment effect ν. Moreover, the weights are given by $\omega_i = 1/\sigma_i^2$.

In general, linear mixed models can be estimated by Maximum Likelihood (ML). However, ML tends to underestimate the variance components. A modified version of ML, known as *restricted maximum likelihood* (REML) is therefore often recommended, since it provides consistent estimates for the variance components (for details see Diggle *et al.*, 2002). ML or REML is possible for the model we described by applying function the **lme** function from `nlme` R-package (Pinheiro *et al.*, 2015).

For the TB dataset, we implement this method with REML in R and the corresponding R-code can be seen in Appendix A.1.2. We obtain estimates for the mean treatment effect -0.75 with a 95% confidence interval of $(-1.11, -0.38)$ and estimated heterogeneity variance of 0.03.

Bayesian inference

The popularity of Bayesian methods has constantly increased in recent years in different research areas including meta-analysis, mainly as a result of advances in computational methods.

Bayesian techniques consider the model parameter θ, as random variable with appropriate *prior distribution* and the data as fixed, as opposed to frequentist techniques where the situation is just the other way around. The *posterior distribution* summarizes the information about θ after observing the data $X = x$. The density function of the posterior distribution, $f(\theta|x)$, is obtained as follows:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta} \tag{1.8}$$

where $f(x|\theta)$ is the likelihood function, $f(\theta)$ is the density function of the prior distribution and the denominator is known as the *marginal likelihood* (Held and Sabanés Bové, 2014c).

Since the marginal likelihood usually does not depend on θ, the following formula can be used for the computation of $f(\theta|x)$:

$$f(\theta|x) \propto f(x|\theta)f(\theta) \tag{1.9}$$

Statistical inference about θ is based solely on the posterior distribution. Suitable point estimates are location parameters, such as the mean, median or mode, of the posterior distribution. Bayesian interval estimates such as *credible intervals* and *highest posterior density intervals* (HPD) are also derived from the posterior distribution. Now, we will proceed with two Bayesian methodology namely empirical Bayes and fully-Bayes.

- **Empirical Bayes**

  Empirical Bayes (EP) methods can be considered as a combination of the Bayesian approach with likelihood techniques. The main distinction of this method is that the EP approach estimate parameters of the prior distribution from data, rather than fixing them based on prior knowledge (Held and Sabanés Bové, 2014b; Davison, 2003).

  When we consider the random effects model, equation (1.5), from empirical Bayes perspective, the corresponding likelihood function and the prior distribution are:

$$f(\hat{\theta}_i|\theta_i) \propto \exp\left\{-\frac{1}{2\sigma_i^2}(\hat{\theta}_i - \theta_i)^2\right\} \quad \text{and} \quad f(\theta_i) \propto \exp\left\{-\frac{1}{2\tau^2}(\theta_i - \nu)^2\right\} \tag{1.10}$$

  As a result, by using formula 1.9, we can derive, analytically, the posterior distribution of each treatment effect θ_i as follows:

$$\begin{aligned} f(\theta_i|\hat{\theta}_i) &\propto f(\hat{\theta}_i|\theta_i)f(\theta_i) \\ &\propto \exp[-\frac{1}{2}\{\frac{1}{\sigma_i^2}(\hat{\theta}_i - \theta_i)^2 + \frac{1}{\tau^2}(\theta_i - \nu)\}^2] \\ &\propto \exp[-\frac{1}{2}(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2})\{\theta_i - (\frac{1}{\sigma_i^2} + \frac{1}{\tau^2})^{-1}(\frac{\hat{\theta}_i}{\sigma_i^2} + \frac{\nu}{\tau^2})\}^2] \end{aligned} \tag{1.11}$$

  Therefore, the posterior distribution is also normally distributed and can be written as follows:

$$\theta_i|\hat{\theta}_i \sim \mathcal{N}(\tilde{\nu}_i, \tilde{\sigma}_i^2) \tag{1.12}$$

  where $\tilde{\sigma}_i^2 = 1/(1/\sigma_i^2 + 1/\tau^2)$ and $\tilde{\nu}_i = \tilde{\sigma}_i^2(\hat{\theta}_i/\sigma_i^2 + \nu/\tau^2)$. However, the application of this formula requires the knowledge of ν and τ^2.

In our random effects model framework, for estimating ν and τ^2, we will use an empirical Bayes method which is achieved by numerical maximization of the marginal likelihood. Initially, for fixed τ^2 the estimate $\hat{\nu}_{ML}(\tau^2)$ of the underlying treatment effect is exactly same as the estimate for ν in the method of moment approach of the random effects model (i.e. the weights are given by $\omega_i = 1/(\sigma_i^2 + \hat{\tau}^2)$). Then, the empirical Bayes estimate for heterogeneity variance, $\hat{\tau}^2_{ML}$, can be estimated by numerically maximizing the profile log-likelihood:

$$\ell_p(\tau^2) = -\frac{1}{2}\sum_{i=1}^{k}[\log(\sigma_i^2 + \tau^2) + \frac{\{\hat{\theta}_i - \hat{\nu}_{ML}(\tau^2)^2\}}{\sigma_i^2 + \tau^2}] \tag{1.13}$$

Empirical Bayes estimates of the individual treatment effects θ_i are finally obtained by plugging the MLEs $\hat{\nu}_{ML}$ and $\hat{\tau}^2_{ML}$ into equation (1.12) in place of the fixed values ν and τ^2.

For the TB dataset, we implement this estimation procedure in R and the corresponding R-code can be seen in Appendix A.1.3. We obtain estimates for the mean treatment effect -0.74 with a 95% confidence interval of $(-1.13, -0.37)$ and estimated heterogeneity variance of 0.3, so very similar results as in the method of moment approach. Figure 1.2 displays 95 % empirical Bayes interval estimates for the individual treatment effects and a 95 % confidence interval based on the profile likelihood of the mean treatment effect ν.

- **Fully-Bayes**

  Fully-Bayes approach, in contrast to the empirical Bayes, assigns a prior distribution to the mean treatment effect ν and the heterogeneity standard deviation τ. For a meta-analysis framework, in most cases, the statistical inferences will be based on *the marginal posterior densities* for ν and τ^2. However, usually those densities do not have closed forms. One way of obtaining those marginal posterior densities is analytically by integrating the joint posterior density over all the other ("nuisance") remaining parameters. But, in many cases especially relatively complicated situations, these integrations can not be achieved analytically. As a result, simulation-based methods and/or numerical or asymptotic approximation procedures are commonly used as estimation techniques in meta-analysis. Before the discussion about the estimation methods, we will give model formulations first.

  Although, in most cases, fully Bayesian approaches to meta-analysis is used for implementation of random effects models, fixed effect models are also possible, since as we mentioned before, fixed effect models are special cases of random effects models. A fully-Bayesian approach to the fixed effect model have implemented a hierarchical model that mirrors this approach as follows (Sutton and Abrams, 2001):

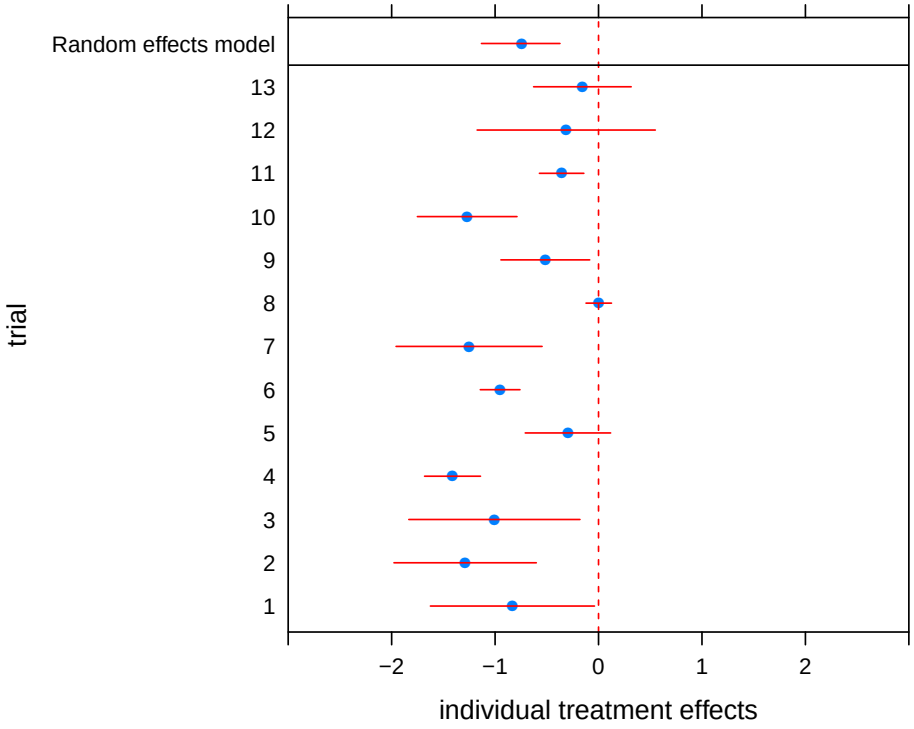

Figure 1.2: 95 % credible intervals for the individual treatment effects θ_i in an empirical Bayes random effects model. Also shown is a 95 % profile likelihood confidence interval for the mean treatment effect ν that is estimated by the empirical Bayes approach ("Random effects model").

$$\begin{aligned} \hat{\theta}_i &\sim \mathcal{N}(\theta, \sigma_i^2) \\ \theta &\sim f(\theta) \end{aligned} \tag{1.14}$$

where f(θ) indicates a prior distribution to be specified. The weights are given by $\omega_i = 1/\sigma_i^2$.

Also, a fully-Bayesian random effects model can be formulated as follows (Sutton and Abrams, 2001):

$$\begin{aligned} \hat{\theta}_i|\theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \mathcal{N}(\nu + \gamma_i, \sigma_i^2) \\ \gamma_i &\sim \mathcal{N}(0, \tau^2) \\ \nu \sim f(\nu) \quad &\text{and} \quad \tau \sim f(\tau) \end{aligned} \tag{1.15}$$

where $f(\nu)$ and $f(\tau)$ indicates prior distributions to be specified. It should be noted that trial specific heterogeneity captured by the random effects $\gamma_i \sim \mathcal{N}(0, \tau^2)$.

Also, the weights are given by $\omega_i = 1/\sigma_i^2$ rather than $\omega_i = 1/(\sigma_i^2 + \hat{\tau}^2)$ as in the empirical Bayes or the method of moment approaches. A very crucial advantage of the Bayesian approach is that it provides posterior distributions of both the mean treatment effect ν and the heterogeneity variance τ^2 besides estimation for individual treatment effects θ_i. Now, we proceed with two estimation methods, namely MCMC and INLA, which can be applied to the fixed effect and random effects models.

MCMC

In recent years, the simulation-based methods for generating values from posterior distributions, in particular the group of methods broadly classified as *Markov Chain Monte Carlo* (MCMC) methods is used quite common as a Bayesian inference technique. It can be said that especially one technique, *Gibbs sampling*, within MCMC class has become a standard choice for modern meta-analysis models. This is because, simulating from typically high dimensional joint posterior densities is often difficult but the posterior conditional distributions are often much easier to sample from. However, our intention is not to give more details about MCMC methods, instead we continue to one last technique, INLA, see Sutton and Abrams (2001) and Higgins *et al.* (2009) for explanations and examples of MCMC methods in the meta-analysis framework.

INLA

The *integrated nested Laplace approximations* (INLA) was very recently proposed by Rue *et al.* (2009) which has become a valid alternative to MCMC. INLA is an approximate Bayesian inference method, so it is a deterministic algorithm rather than simulation-based such as MCMC. The main advantage of INLA is that it provides accurate results in shorter computing time. Furthermore, there is no need to examine convergence of samples as in MCMC. The `INLA` R package, hereafter referred as `r-inla`, provides an interface for R programming language to INLA (a free-standing programme) so that models can be fitted using standard R commands. The `r-inla` package is available on INLA website (http://www.r-inla.org/).

In Sauter and Held (2015), it was shown that INLA approach is suitable for an estimation method of meta-analysis models both for fixed effect and random effects models. INLA approach by using `r-inla` is our main focus in the thesis, from now on, mostly, we will use this Bayesian method for estimation.

A crucial component of any Bayesian inference is the prior specification. When there is no extra information about the parameters besides the available data, the prior distributions can be specified by the "just proper uninformative" distributions (Spiegelhalter *et al.*, 2004). By taking from Sauter and Held (2015), we prefer to set the prior distributions for treatment effect to $\theta \sim \mathcal{N}(0, 1000)$ (or mean treatment effect to $\nu \sim \mathcal{N}(0, 1000)$) and heterogeneity standard deviation to $\tau \sim \mathcal{U}(0, 10)$.

For the thesis, mainly adapting from Sauter and Held (2015), we wrote an R function, `meta.inla`, in order to fit different meta-analysis models described in this chapter. The `meta.inla` function is actually a wrapper for **inla** function from `r-inla` package. In order to use this function, dataset should be brought to a suitable format which can be achieved by `creatINLAdat.dir` function. Both functions are provided in Appendix A.1.4.

For the TB dataset, after the data preparation procedure, the fixed effect model can be implemented in R by specifying the argument `mod = "FE"` within **meta.inla** command as follows:

```
> TB.datINLA <- creatINLAdat.dir(ntrt = TB$TRT, nctrl = TB$CON,
+     ptrt = TB$TRTTB, pctrl = TB$CONTB)
> library(INLA)
> inla.fe.tb <- meta.inla(TB.datINLA, meanf = 0, varf = 1000, mod = "FE")
> print(inla.fe.tb)
Call: meta.inla(datINLA = TB.datINLA, meanf = 0, varf = 1000, mod = "FE")
Meta analysis using INLA
Posterior mean of treatment effect = -0.44    95% CrI ( -0.52, -0.35 )
```

For random effects model, to have a uniform prior for heterogeneity standard deviation, we used a user-specified prior since the uniform prior is not defined under the `r-inla`. This is achieved in `r-inla`, by handing over a table with the prior density evaluated at an appropriate grid. For the details, see Section 1.1 of Supplementary Material of Sauter and Held (2015). Then, by applying **meta.inla** function with the argument `mod = "RE"`, the random effects model can be implemented. We obtain posterior mean ν of -0.75 (posterior median is -0.75) and posterior mean for τ^2 of 0.48 (posterior median is 0.41). The corresponding R-code can be seen in Appendix A.1.5.

As a summary of all estimation approaches which we introduced and implemented until here, Table 1.3 demonstrates estimates for ν (or θ) and τ^2 with their corresponding interval estimates (if they are available) for each approach. Before giving discussions about regression approaches within meta-analysis context, we describe a different type of modelling approach (trial-arm level data) which is crucial especially for meta-analysis models that will be introduced in the following chapters.

1.3.5 Trial-arm level data

It can be said that all these methods we presented, suffer some theoretical disadvantages. Firstly, the assumption of normally distributed log-odds ratios may be inadequate for some situations, for example when result of some studies based on small numbers of events. Especially, if there is a zero count in the contingency table of a trial, then the corresponding odds ratio is not finite. Also none of the meth-

Table 1.3: Results of meta-analysis of TB dataset using a classical fixed effect (inverse-variance weighted), a Bayesian (INLA) fixed effect, a classical random effects by method of moments approach (MOM), a classical random effects in the likelihood approach by REML (REML), an empirical Bayes method (EB) and a Bayesian (INLA) random effects model.

	Fixed effect		**Random effects**			
	Classical	INLA	MOM	REML	EB	INLA
Treatment effect (ν)						
Overall/posterior mean	-0.4361	-0.4368	-0.7474	-0.7486	-0.7420	-0.7475
Lower b.(95%CI/CrI)	-0.5190	-0.5198	-1.1242	-1.1136	-1.1318	-1.1633
Upper b.(95%CI/CrI)	-0.3533	-0.3537	-0.3706	-0.3836	-0.3728	-0.3429
Heterogeneity variance (τ^2)						
Estimated/posterior mean			0.3663	0.0301	0.3025	0.4832
Lower b.(95%CI/CrI)						0.1449
Upper b.(95%CI/CrI)						1.2564

ods takes into account of the fact that σ_i^2 are estimated from the data rather than known (Thompson and Sharp, 1999). These problems are overcome by specifying a logistic regression model with a fully-Bayesian approach which does not require the assumption of normality for the individual treatment effects. With this model, *trial-arm level* summaries are the natural inputs to different inference methods, i.e. they are the data (Dias *et al.*, 2010). This is different than what we were using in our previous models which are called *summary level data*.

Now, we describe models which use the binomial structure of the data directly. The number of events y_{i1} and number of patients n_{i1} is observed in the control arm for each trial $i = 1, 2, \ldots, k$. And correspondingly y_{i2} and n_{i2} in the treatment arm. The number of events in each arm in each trial can be specified as distributed binomially, i.e. $y_{i1} \sim \text{Bin}(\pi_{i1}, n_{i1})$ and $y_{i2} \sim \text{Bin}(\pi_{i2}, n_{i2})$. The log-odds ratio ν can now be modeled as with logistic regression as:

$$\begin{aligned} \text{logit}(\pi_{i1}) &= a_i \\ \text{logit}(\pi_{i2}) &= a_i + \nu + \gamma_i \end{aligned} \tag{1.16}$$

where the treatment effect a_i of control arm in trial i is a nuisance parameter and the main interest is in the log odds ratio ν. The trial specific heterogeneity captured by the random effects $\gamma_i \sim \mathcal{N}(0, \tau^2)$. If $\tau^2 = 0$, we obtain a fixed effect model. Moreover, no weights are used in this case unlike the models for summary level data.

We use only a Bayesian approach with INLA in order to fit the TB dataset as a trial-arm level. Firstly, the data should be brought into a suitable format, namely one-arm-per-row data. This can be achieved by using **creatINLADAT.dir** function. Then, the fixed effect model and random effects model using trial-arm level can be implemented via **meta.inla** function with the argument `type = "trial-arm"`. Table 1.4 demonstrates the results of this procedure with fixed effect model and random effects model, the results are similar which we obtained in INLA approach with summary-level data. The corresponding R-code can be seen in Appendix A.1.5. However, for the sake of convenience we are not giving more details about this method, we will focus on such models in following chapters.

Table 1.4: Results of meta-analysis of TB dataset as a trial-arm level data using fixed effect model and random effects model by logistic regression method in fully Bayes approach with INLA.

	Fixed effect	**Random effects**
Treatment effect (ν)		
Overall/posterior mean	-0.4784	-0.7610
Lower b.(95%CI/CrI)	-0.5597	-1.1801
Upper b.(95%CI/CrI)	-0.3975	-0.3542
Heterogeneity variance (τ^2)		
Estimated/posterior mean		0.4967
Lower b.(95%CI/CrI)		0.1496
Upper b.(95%CI/CrI)		1.2920

1.4 Meta-regression

Random effects models take into account the heterogeneity between different studies as we described, however they do not provide a method of exploring and potentially of explaining sources of heterogeneity. In order to explore the between-study variation or possible reasons why study results vary systematically, regression type models, named *meta-regression*, have been used (Lau *et al.*, 1998). In contrast to meta-analysis methods which we discussed so far, meta-regression models examine the associations between the characteristics of the trials involved and treatment effects. It is important to mention that we use the term meta-regression to indicate the use of summary level or trial-arm level covariates, as distinct from regression analyses that are possible when individual patient data on outcomes and covariates are available.

On the other hand, even if we only use randomized trials data, the study of covariates is inherently observational, since it is not possible to randomize patients to one covariate. As a consequence, meta-regression has many difficulties of interpretation and inference

which attach to non randomized trials such as confounding, correlation between covariates and, very important, the inability to infer causality from association.

There are various statistical techniques for meta-regression. Those techniques differ in a number of aspects; for example whether the underlying model is assumed as fixed effect or random effects model, whether estimation procedure is a classical approach or a Bayesian approach (Thompson and Sharp, 1999). In this section, firstly weighted-least squares and likelihood with REML approaches of random effects meta-regression within classical perspective will be discussed, then a Bayesian approach with INLA methodology will be shown and finally we will describe meta-regression using trial-arm level dataset.

1.4.1 Classical inference

Weighted-least squares approach

Firstly, we introduce a random effects meta-regression model by a weighted-least squares approach within classical inference. The log-odds ratios of each trial is assumed to follow a normal distribution and the regression uses weighted least squares approach to take into account the variance of the log-odds ratio estimate in each trial. The model formulation which is an extension of equation (1.2) can be shown as follows:

$$\hat{\theta}_i \sim \mathcal{N}(\nu + x_i \cdot \beta, \sigma_i^2 + \tau^2) \tag{1.17}$$

where x_i is a study-level covariate, β represents the change in $\hat{\theta}_i$ per unit of change in covariate x_i and ν is the log-odds ratio at $x_i = 0$ (Thompson and Sharp, 1999). Also, between-study variance τ^2 is added to within-study variance σ_i^2 for incorporating heterogeneity between trials. Maximum likelihood estimates of a and β can be obtained by least square regression of $\hat{\theta}_i$ on x_i with weights $w_i = 1/(\sigma_i^2 + \hat{\tau^2})$. In order to fit this model, two-step approach is proposed (Raudenbush, 2009), first τ^2 is estimated with one of the available estimators, then undertake the linear model via weighted least squares.

This methodology is implemented in R package `metafor` (Viechtbauer, 2010) and we will utilize the functionality from this package to analyse the TB dataset with covariates **Year** and **Latitude** by using method of moment estimator for τ^2 which we gave in Section 1.3.4 (see Schwarzer *et al.*, 2015 for different meta-analysis R packages overview).

Firstly we are shifting the covariates to a meaningful center, since this makes easier to interpret the results. We set this center as characteristics of study number 5 which has the minimum latitude available among trials included. That means, for **Latitude** as 13 and for **Year** as 1973. Then, by applying function **rma.uni** to TB dataset, meta-regression can be fitted in R. The corresponding R-code can be seen in Appendix A.1.6.

As a result, we obtain the estimate of the intercept as -0.13 (95 % CI -0.46, 0.19) which have an interpretation as the expected treatment effect of a study with **Latitude** of 13 and **Year** of 1973. Also, we got the estimate of regression coefficient for **Latitude**,

β_{Lat}, -0.03 (95 % CI -0.05, -0.01). β_{Lat} means the expected increase in treatment effect with every one point increase in **Latitude**. And similarly, β_{Year} is 0 (95 % CI -0.02, 0.03) which is the expected increase in treatment effect with every one point increase in **Year**. Table 1.5 demonstrates these results with estimated heterogeneity variance.

Mixed model approach

Now, we proceed with a different approach to random effects meta-regression which is achieved by using a linear mixed model. That procedure is an extension of equation (1.7) and model formula can be obtained by including a study-level covariate, denoted x_i, as follows:

$$\begin{aligned} \hat{\theta}_i|\theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &= \nu + x_i \cdot \beta + \gamma_i + \epsilon_i \end{aligned} \tag{1.18}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ is sampling error within each trial, $\gamma_i \sim \mathcal{N}(0, \tau^2)$ is random effect for trial specific heterogeneity and ν is the log odds ratio at $x_i = 0$. The weights are given by $\omega_i = 1/\sigma_i^2$. The maximum likelihood estimation or restricted maximum likelihood (REML) estimation can be achieved as in equation (1.7) but in this case simply adding covariates to the model formula when applying the **lme** function in R.

For the TB dataset, we implement this method with REML estimation method in R by including **Year** and **Latitude** (again as centered) as a study level covarites. The corresponding R-code can be seen in Appendix A.1.6. Table 1.5 shows the results which are very similar to the weighted least squares approach except the heterogeneity variance. Figure 1.3 displays a visual presentation of the results.

As we pointed out, meta-regression investigates whether particular covariates explain any of the heterogeneity of treatment effects between studies. It is not reasonable to assume that all of the heterogeneity is explained. Therefore meta-regression with a random effect analysis is more appropriate than a fixed effect one (Thompson and Higgins, 2002). For this reason, we are not presenting any fixed effect meta-regression model.

1.4.2 Bayesian inference

A fully-Bayesian approach to random effects meta-regression model can be obtained by including covariates in a Bayesian meta-analysis model, equation (1.15), is as follows:

$$\begin{aligned} \hat{\theta}_i|\theta_i &\sim \mathcal{N}(\theta_i, \sigma_i^2) \\ \theta_i &= \nu + x_i \cdot \beta + \gamma_i + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, \sigma_i^2) \\ \gamma_i &\sim \mathcal{N}(0, \tau^2) \end{aligned} \tag{1.19}$$

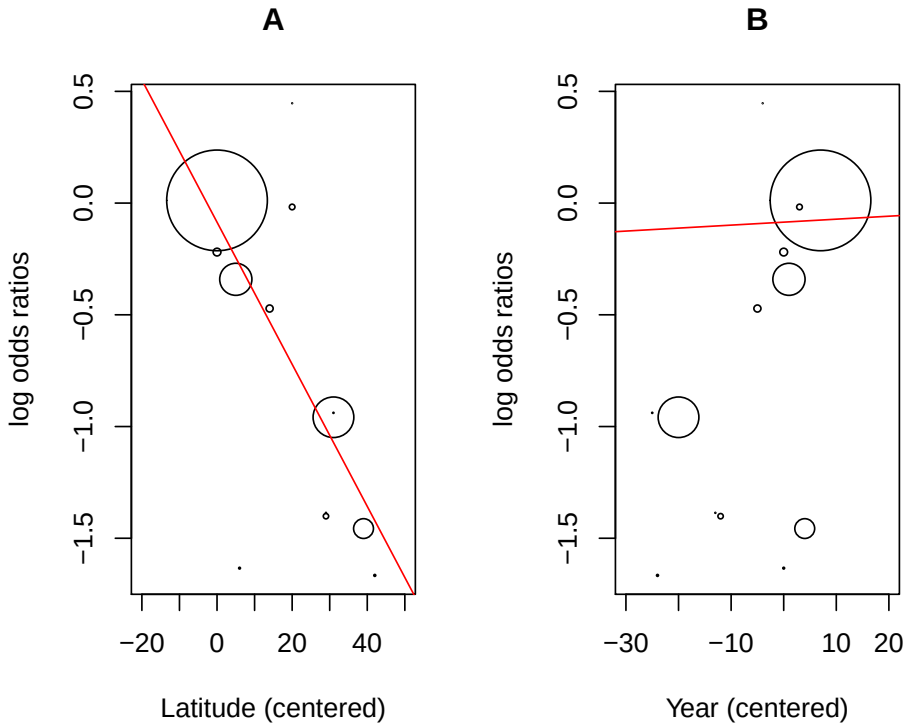

Figure 1.3: Log odds ratios of the TB dataset, including 13 trials, according to in (A) the geographic latitude of the place where the trial was undertaken which is centered at 13 degree and in (B) the year of publication which is centered at 1973. The circle corresponding to each trial has area proportional to the the weights w_i. The superimposed lines are obtained by mixed model approach of random effects meta-regression using an REML estimate.

Similar to the mixed model approach, the weights are given by $\omega_i = 1/\sigma_i^2$ rather than $\omega_i = 1/(\sigma_i^2 + \hat{\tau}^2)$.

We prefer to set the prior distributions for all covariates to a normal distribution with mean zero and variance 1000. After data preparation using **creatINLAdat.dir** meta-regression using random effects model can be implemented using **meta.inla** function with the argument `mreg = "TRUE"` as follows:

```
> TB.mreg.datINLA <- creatINLAdat.dir(ntrt = TB$TRT, nctrl = TB$CON, ptrt = TB$TRTTB,
+            pctrl = TB$CONTB, cov1 = TB$Latitude_centered, cov2 = TB$Year_centered)
> inla.mreg.re.tb <- meta.inla(TB.mreg.datINLA, meanf = 0, varf = 1000,
+                          ul = 10, mod = "RE", mreg = "TRUE")
> print(inla.mreg.re.tb, digits = 3)
Call: meta.inla(datINLA = TB.mreg.datINLA, meanf = 0, varf = 1000,
    ul = 10, mod = "RE", mreg = "TRUE")
Meta regression using INLA
Intercept = -0.163    95% CrI ( -0.696, 0.322 )
Latitude = -0.028    95% CrI ( -0.053, 0 )
```

```
Year = 0.006    95% CrI ( -0.029, 0.044 )
Posterior mean of heterogeneity variance = 0.12    95% CrI ( 0.015, 0.764 )
```

We obtain similar results with the weighted least squares approach, see Table 1.5 for the results.

Table 1.5: Results of random effects meta-regression analysis of TB dataset: Weighted-least squares approach (WLSQ), mixed model approach with REML (REML), fully Bayesian approach by INLA with summary level dataset (INLA) and INLA approach with trial-arm level dataset (INLA-ARM). **Latitude** and **Year** are included as covariates to the models.

	WLSQ	**REML**	**INLA**	**INLA-ARM**
Intercept				
Estimated/posterior mean	-0.1316	-0.0854	-0.1631	-0.1618
Lower b.(95%CI/CrI)	-0.4561	-0.3687	-0.6960	-0.6962
Upper b.(95%CI/CrI)	0.1930	0.1980	0.3218	0.3226
Latitude				
Estimated/posterior mean	-0.0295	-0.0318	-0.0278	-0.0281
Lower b.(95%CI/CrI)	-0.0461	-0.0462	-0.0527	-0.0529
Upper b.(95%CI/CrI)	-0.0129	-0.0174	-0.0003	-0.0006
Year				
Estimated/posterior mean	0.0046	0.0013	0.0063	0.0069
Lower b.(95%CI/CrI)	-0.0197	-0.0207	-0.0291	-0.0282
Upper b.(95%CI/CrI)	0.0288	0.0233	0.0442	0.0453
Heterogeneity variance				
Estimated/posterior mean	0.0667	0.0002	0.1203	0.1263
Lower b.(95%CI/CrI)			0.0147	0.0143
Upper b.(95%CI/CrI)			0.7639	0.7647

1.4.3 Trial-arm level data

The model formulation of meta-regression by using summary level data has same disadvantages as in meta-analysis context (without regression), hence meta-regression random effects model with logistic regression by using trial-arm level data is preferable, in principle. A Bayesian model formulation for this type of meta-regression can be obtained by including the study level covariate x_i to equation (1.19) (Thompson and Sharp, 1999). Therefore, the number of events is distributed as $y_{i1} \sim \text{Bin}(\pi_{i1}, n_{i1})$ and $y_{i2} \sim \text{Bin}(\pi_{i2}, n_{i2})$, produces the following model:

$$\begin{aligned} \text{logit}(\pi_{i1}) &= a_i \\ \text{logit}(\pi_{i2}) &= a_i + \nu + x_i \cdot \beta + \gamma_i \end{aligned} \tag{1.20}$$

Similar to the model for summary level data, trial-specific heterogeneity is captured by $\gamma_i \sim \mathcal{N}(0, \tau^2)$. A random effects meta-regression using a trial-arm level can be fitted with **meta.inla** using the arguments `mod = "RE"`, `type = "trial-arm"` and `mreg = "TRUE"`. The corresponding R-code can be seen in Appendix A.1.7. Table 1.5 shows the results of this methodology.

R version and packages used to generate this chapter:
R version: R version 3.2.3 (2015-12-10)
Base packages: splines, stats, graphics, grDevices, utils, datasets, methods, base
Other packages: metafor, INLA, Matrix, sp, nlme, lattice, xtable
Versions of other packages (respectively): 1.9.8, 0.0.1443538834, 1.2.3, 1.2.2, 3.1.124, 0.20.33, 1.8.2
This document was generated on February 23, 2016 at 13:24.

Chapter 2

Network meta-analysis

2.1 Introduction

One of the most reliable ways of comparing two treatments is the direct comparison of randomized trials as in the TB dataset in Chapter 1 (in that case, for two treatments, we considered treatment vs control). In order to analyse the results of collected evidence from different studies which include only direct comparisons or, shortly, for meta-analysis of direct comparisons, there are different approaches as we discussed in Chapter 1. This type of meta-analysis can be called *pairwise meta-analysis* or *conventional meta-analysis* or *head-to-head meta-analysis* (Salanti, 2012). However, in many areas, available trials may not have directly compared the specific treatments of interest. This situation can stem from different reasons. For example, if there may be a class of several treatments, each of which has been studied placebo-controlled randomized trials. If there are no trials or very few trials in which the drugs have been directly compared with each other, then by a pairwise meta-analysis, the comparison between two treatments from this class is not possible, Therefore, there is a need for a broader and inclusive view of the available evidence rather than pairwise meta-analysis (Salanti, 2012).

2.1.1 Direct vs. indirect estimates

Before the explanation of this broader approach or so-called *network meta-analysis* (NMA), we will use some simple examples to introduce concepts of *direct* and *indirect estimates*. As a first example, say Network 1, assume that there are two type of studies namely AC studies which compare treatment A and control C and BC studies which compare treatment B and control C (the formal definition of a *network* will be given in Section 2.2). Also, as Network 2 suppose that there are AC, BC and AB studies. We can represent the available treatment comparisons using a simple graph. Figure 2.1 shows such graphical representations of these two networks.

Firstly, when we consider Network 1, AC studies provide a *direct estimate* of the difference of treatment effects of A and C measured on some scale, say log odds ratio as

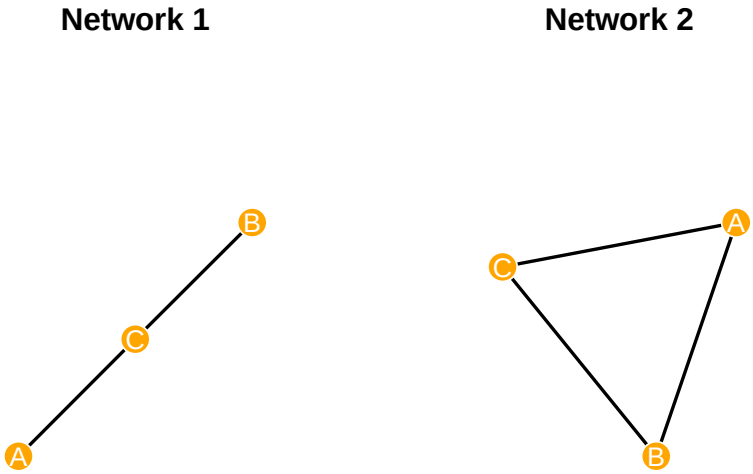

Figure 2.1: Graphical representation of two networks with three treatments (A, B and C). Network 1 includes AC and BC studies. Network 2 includes AC, BC and AB studies. Lines indicate we have data from one or more studies comparing the two treatments.

in examples in Chapter 1. We denote this relative treatment effect as d_{AC}^{Dir} (where the superscript denotes the direct estimate). Other studies (BC studies) may provide information on the direct comparison between treatment B and the same control C; denoted d_{BC}^{Dir}. Also, AC and BC studies provide an *indirect estimate* for the comparison of A and B from the relative treatment effect of A-C and B-C as follows (Schwarzer *et al.*, 2015):

$$d_{AB}^{Ind} = d_{AC}^{Dir} - d_{BC}^{Dir} \tag{2.1}$$

Secondly, when we consider Network 2, there are d_{AC}^{Dir}, d_{BC}^{Dir} and d_{AB}^{Ind} as in Network 1. Additionally, there is direct evidence from studies comparing A and B (AB studies), denoted by d_{AB}^{Dir}. In summary, in a simple example like Network 2, for every comparison between two treatments, one can estimate direct and indirect estimates.

Now, consider a slightly more complex example, say Network 3 which has an additional treatment D with AD and BD studies. Figure 2.2 shows this network. When we examine this network, the treatments, called *nodes* in graph theory, which are joined with a line, called *edge* in graph theory, correspond to those for which direct evidence is available. Thus, from Figure 2.2 we see that direct evidence is available for all comparisons except

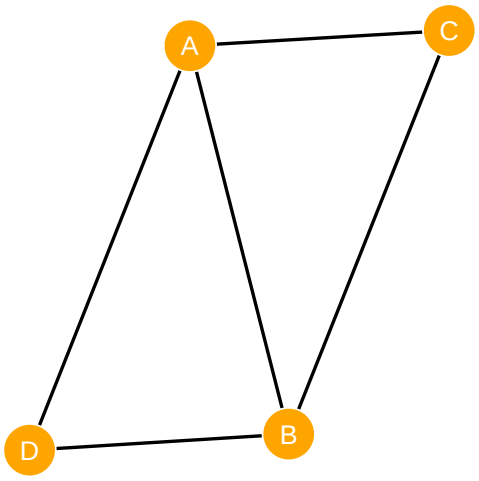

Figure 2.2: Graphical representation of Network 3 which is a network with four treatments, A, B, C and D. Lines indicate we have data from one or more studies comparing the two treatments.

between C and D which must be estimated indirectly. Moreover, indirect estimates, for instance d_{AB}^{Ind}, can be derived via two possible ways namely by using AD and BD studies in addition to by using AC and BC studies.

At this point, we can give the definition of NMA. NMA or *mixed treatment comparisons* (MTC) is used for combining direct and indirect estimates across a network of randomized trials to infer about the relative treatment effect of multiple treatments (Salanti, 2012). For simplicity, we begin by considering trials consisting of only two-arm trials. On the other hand, there is a different kind of trial named *multi-arm trial* which compares several treatments against a common control arm (Jaki, 2015). The extension to multi-arm trials in NMA framework is possible and will be discussed later. As a side note, network meta-analysis is a generalization of pairwise meta-analysis.

2.2 Some terminology in network meta-analysis

Network meta-analysis is a sufficiently new research area and terminology differs between authors and continues to evolve (Puhan *et al.*, 2014). Now we will present some terms used in this thesis.

A *network* is a collection of trials of alternative treatments for a common clinical

condition (such as disease) that allow, through direct and indirect estimates, calculation of the relative treatment effects of all treatment versus one another on a particular outcome (Puhan *et al.*, 2014).

2.2.1 Transitivity

In NMA context, a crucial concept is the assumption of *transitivity*. That is an underlying assumption when d_{AB}^{Ind} is calculated —one can learn about A versus B via C. In other words, an indirect comparison validly estimates the unobserved direct comparison (Salanti, 2012). Therefore, the violation of the assumption of transitivity, *intransitivity*, refers to differences in study characteristics that may modify treatment effect in the direct comparisons (such as d_{AC}^{Dir} and d_{BC}^{Dir}) that form the basis for the indirect estimate of d_{AB}^{Ind}, and thus bias the indirect estimate of A versus B (Puhan *et al.*, 2014). However, this cannot be tested statistically, but its plausibility can be evaluated conceptually and epidemiologically. Therefore, this is not our focus in the thesis but see Baker and Kramer (2002) for a discussion about this topic.

2.2.2 Consistency

Consistency in a network means that there is no discrepancy between direct and indirect estimates (Sauter and Held, 2015). The assumption of consistency is linked to the assumption of transitivity, as the former is the extension of the latter. Mathematically speaking, the assumption of consistency is reflected by equation (2.1) for a network includes treatment A, B and C. When we consider such network, if there is no direct evidence for the relative treatment effect of A vs B (no d_{AB}^{Dir}) or in other words no available comparisons of A and B, then the consistency assumption of this comparison reduces to transitivity. If there is direct evidence to estimate d_{AB} (left part of the equation), consistency claims that the two pieces of evidence give the same result. Unlike the transitivity, statistical methods can be used to evaluate consistency (Salanti, 2012). In the next section, we will discuss the assessment of *network inconsistency* which is the examination of the possibility that consistency restrictions are not fulfilled. As a side note, this term has been referred as *incoherence* by Lumley (2002).

The heterogeneity is already introduced in Chapter 1. With heterogeneity, we mean differences between trials comparing the same treatments but being different in terms of trial-specific features, e.g. differences between study-populations. We will later discuss the statistical model named *consistency model* which incorporate heterogeneity by assuming consistency in the network and also the model called *inconsistency model* which incorporate both heterogeneity and inconsistency in the network.

2.2.3 Basic and functional contrasts

Here, we describe the parametrization of a network and introduce some terms from graph theory by mainly adapting from Lu and Ades (2006). The parametrization is crucial especially for defining statistical models for NMA. In a network meta-analysis the primary interest is, usually, to compare a number of treatments to a *baseline treatment*. This can be placebo, usual care, no treatment, or a well-established standard treatment (Schwarzer *et al.*, 2015). Consider a network, say Network 4, with four treatments (A, B, C, D) such that every comparisons between any two treatments are available. Figure 2.3 illustrates this network.

Here, if treatment C, control, is considered the baseline treatment for relative treatment effects, then three relative treatment effect parameters d_{XC} for $X = A, B$ or D are the *basic contrasts*. And the other three (d_{AB}, d_{DB} and d_{AD}) are *functional contrasts* that can be represented as functions of basic contrasts through the following linear relations:

$$\begin{aligned} d_{AB} &= d_{AC} - d_{BC} \\ d_{DB} &= d_{DC} - d_{BC} \\ d_{AD} &= d_{AC} - d_{DC} \end{aligned} \tag{2.2}$$

Functional contrasts can be written in terms of basic contrasts under the assumption of consistency. Basic contrasts can be chosen in different ways based on NMA modelling approach and will be discussed later. Let $\mathbf{d}_b$ be the vector of basic contrasts and $\mathbf{d}_f$ be the vector of functional contrasts. The number of basic contrasts equal to the number of treatments in the network, say K, minus one, $\#\ (\mathbf{d}_b) = K - 1$. Therefore, the number of functional contrasts equal to the number of available treatment comparisons, say T, minus the number of basic contrasts, $\#\ (\mathbf{d}_f) = T - K - 1$. Each consistency relation corresponds to a *cycle*, a path that starts and ends at the same node, of edges in the network graph. In the above example, the corresponding cycle of the relation of $d_{AB} = d_{AC} - d_{BC}$ which is formed by AB, AC and BC edges (thick lines) in Network 4 can be seen in Figure 2.3. In the literature, cycle and loop are used interchangeably for example in Salanti (2012) or in Sauter and Held (2015), however we prefer to use cycle instead of loop, since, in graph theory, loop is defined as an edge that connects a node to itself (West *et al.*, 2001). This definition is not coincided with the meaning of loop in the context of network meta-analysis, for this reason, in our opinion the usage of loop may lead to confusion.

A *3-cycle* means a cycle with three nodes —treatments in NMA context (West *et al.*, 2001). A *spanning tree* in a network graph is a connected subgraph consisting of all nodes but containing no cycles. For Network 4, one spanning tree is formed by AB, AC and AD edges (dashed lines) in Figure 2.3. Therefore, if one new edge is added to a spanning tree, the new graph represents one functional contrasts and forms one cycle. By the time that three new edges are added to the spanning tree formed by AB, AC, and AD in Figure 2.3,

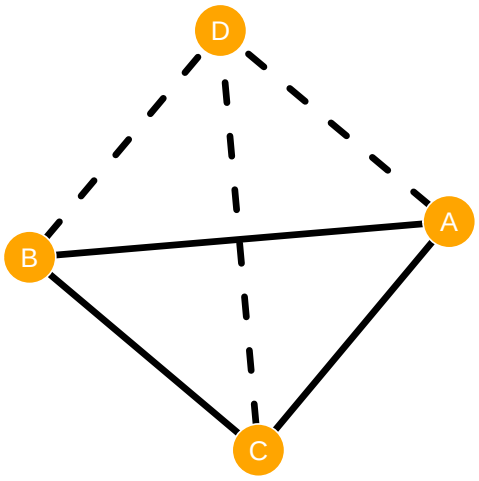

Figure 2.3: Graphical representation of Network 4. Thick lines correspond to a 3-cycle formed by AB, AC and BC edges in the network graph. Dashed lines correspond to a spanning tree formed by AD, CD, BD edges.

a total of seven cycles have been created. However, from all seven cycles, only three cycles are *independent* in the sense that if we know that the relations in these three cycles are consistent, that means all seven are consistent (see Section 2.2 and 2.3 in Lu and Ades, 2006).

2.3 Statistical models for network meta-analysis

Our general strategy to formulate the models for NMA can be summarized as writing down a hierarchical model containing components for sampling variability, heterogeneity and inconsistency. In Section 2.3.1, we introduce a NMA model for summary level data which is a Bayesian linear mixed model. Section 2.3.2 discusses first a NMA model for trial-arm level data with a binomial outcome and then extends this model in order to account for multi-arm trials which is a Bayesian generalized linear mixed model with logit link function.

2.3.1 Summary level approach

This model is introduced by Lumley (2002). Actually we used this approach before in Chapter 1 (more precisely see equation (1.7)). Now we will give a more general model formulation of this approach so that it can be used for fitting NMA models. Consider treatment effects parameters θ_{ijk} comparing treatment j with k in trial i with corresponding within-trial variance σ^2_{ijk} for several independent two-arm trials $i = \{1, 2, \ldots, S\}$. The treatment pair $k, j \in \{1, \ldots, T\}$ compared in trial i is one among $T(T-1)/2$ possible combinations. The log-odds ratio θ_{ijk} is assumed to follow a normal distribution and is modelled as

$$\begin{aligned} \theta_{ijk} &\sim \mathcal{N}(d_{jk} + \gamma_{ijk} + \xi_{jk}, \sigma^2_{ijk}) \\ \gamma_{ijk} &\sim \mathcal{N}(0, \tau^2) \\ \xi_{jk} &\sim \mathcal{N}(0, \kappa^2) \end{aligned} \tag{2.3}$$

The relative treatment effect d_{jk} is the difference between the treatment effects d_j and d_k such that $d_{jk} = d_j - d_k$. Consider the comparison of treatment j with k in trial i where we now introduce a random effect γ_{ijk}. This heterogeneity random effects capture differences between trials comparing the same treatments but being different in terms of trial-specific features. Thus, it has a similar interpretation as in the pairwise meta-analysis context.

In order to make the model identifiable, we need to fix the treatment effect of some arbitrary baseline treatment at zero. By consequence we only need $T-1$ parameters to fully describe the model with its network structure. If there exists a treatment such that all possible comparisons between that treatment and every other treatment are available, then we call it *global baseline*. For illustrative reasons, consider a simple fully connected network with three treatments (1, 2 and 3). In this network, if treatment 1 is chosen as global baseline, then basic contrasts are $\mathbf{d}_b = (d_{12}, d_{13})^T$ and the only functional contrast is $d_f = d_{23}$. Under the assumption of consistency, we have $d_{23} = d_{13} - d_{12}$, i. e. $\mathbf{d}_f = \mathbf{F}^T \mathbf{d}_b$ where $\mathbf{F}^T = (1, -1)^T$ (Lu and Ades, 2006). In general, if there exists a global baseline treatment, say treatment 1, then we have basic contrasts $\mathbf{d}_b = (d_{12}, d_{13}, \ldots, d_{1T})^T$, the treatment effects relative to the global baseline treatment. Based on $\mathbf{d}_b$ we can fully describe the network structure or, more precisely, $\mathbf{d}_b$ forms a spanning tree in the network graph.

In the model, there are additional random effects, inconsistency random effects, ξ_{jk} which capture the network inconsistency. For instance, for the simple network above, this model introduces three random effects, namely ξ_{12}, ξ_{13} and ξ_{23}. We assume that $\xi_{jk} \sim \mathcal{N}(0, \kappa^2)$. The inconsistency variance κ^2 is a measure for the degree of inconsistency in the network. As a side note, when assume $\kappa^2 = 0$ in the model, it corresponds to a consistency model.

2.3.2 Trial-arm level approach

The model is discussed in this section has been introduced by Lu and Ades (2006). In Chapter 1 we have already discussed trial-arm level and summary level datasets. A NMA model is possible by using trial-arm level data that is a similar approach which we introduced in Section 1.3.5 (model formulation of equation (1.16)). Here, firstly we give a model formulation for NMA models including two-arm trials only, then a model which account for multi-arm trials will be introduced.

Each trial $i = \{1, 2, \ldots, S\}$ has two treatment arms namely $t_1(i)$ and $t_2(i)$. The first treatment $j = t_1(i)$ is chosen as baseline treatment and compared with the other treatment $k = t_2(i)$. For each trial i and baseline treatment j the number of events y_{ij} and number of patients n_{ij} are observed. Correspondingly, also for treatment k, y_{ik} and n_{ik} are observed. The number of events is (conditionally) independent for each trial-arm and follows a binomial distribution, i.e. $y_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij})$ and $y_{ik} \sim \text{Bin}(n_{ik}, \pi_{ik})$. The log-odds ratio d_{jk} of baseline treatment j vs. treatment k can now be modelled with logistic regression as:

$$\begin{aligned} \text{logit}(\pi_{ij}) &= a_{ij} \\ \text{logit}(\pi_{ik}) &= a_{ij} + d_{jk} + \gamma_{ijk} + \xi_{jkl} \end{aligned} \tag{2.4}$$

The treatment effect a_{ij} of baseline treatment j in trial i is a nuisance parameter and the main interest is in the log-odds ratio d_{jk}. Similar as in Section 2.3.1, possible trial-specific heterogeneity is captured by the random effects $\gamma_{ijk} \sim \mathcal{N}(0, \tau^2)$.

For inconsistency, a cycle-specific approach is proposed by Lu and Ades (2006) which is adding a random effect for every independent 3-cycle (see also Dias *et al.*, 2010; Sauter and Held, 2015). Therefore, here consistency is a property of a cycle of a network graph, which in the case of three treatments is simply a triangle, rather than a pairwise comparison. Consider a 3-cycle with treatments j, k, l, where we now introduce a random effect $\xi_{jkl} \sim \mathcal{N}(0, \kappa^2)$. Here, in order to account for inconsistency, we relax the consistency relation $d_{lk} = d_{lj} - d_{jk}$ to $d_{lk} = d_{lj} - d_{jk} + \xi_{jkl}$. The number of cycle-specific inconsistency random effects is called the inconsistency degrees of freedom (ICDF). When there is no multi-arm trial in the network, the ICDF is equal to number of functional contrasts (ICDF = # (d_f)). By conducting a trial-arm level approach with a Bayesian method, the posterior distribution of ξ_{jkl} reflects the extent of inconsistency in the cycle with j, k and l treatments of the network.

In a trial-arm level data approach, the parametrization is different compared to a summary level Here, it is not necessary to use a global baseline treatment to parametrize the network. Any subset of relative treatment effect parameters can be chosen as basic contrasts $\mathbf{d}_b$ as long as their corresponding edges form a spanning tree in the network graph. As in general case, the remaining functional contrasts $\mathbf{d}_f$ can be described as linear combinations of $\mathbf{d}_b$.

Multi-arm trials

Often, multi-arm trials, trials with more than two treatment arms, are included in a network meta-analysis. Nowadays, in a variety of disease areas, a number of multi-arm trials are being run (Jaki, 2015). Furthermore, according to Parmar *et al.* (2014), because of the advantages offered by multi-arm trials compared to two-arm trials, more multi-arm randomized trials are needed. One important advantage of multi-arm trials compared to separate two-arm trials to assess several treatments is that the multi-arm design is quicker and cheaper.

By adapting from Sauter and Held (2015), we will introduce a statistical model which accounts for multi-arm trials. Here, each trial i can have more than two treatment arms namely $t_1(i), \ldots, t_{K_i}(i)$. Again, the first treatment $j = t_1(i)$ is chosen as baseline treatment and compared with the remaining treatments $k = t_2(i), \ldots, t_{K_i}(i)$. As in the case of only two-arm trials, we assume that $y_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij})$, $y_{ik} \sim \text{Bin}(n_{ik}, \pi_{ik})$ and d_{jk} is modelled with the equation (2.4).

In the model that accounts only for two-arm trials, it is assumed that $\gamma_{ijk} \sim \mathcal{N}(0, \tau^2)$. Therefore, for a fixed i, $\gamma_i = \gamma_{ijk}$, since there are only two arms to compare. However with the presence of multi-arm trials, there are several treatments that are compared to baseline treatment within one trial.

Consider a simple situation where a trial i includes treatments 1, 2 and 3 and treatment 1 is the baseline treatment. Now heterogeneity random effect, $\boldsymbol{\gamma}_i = (\gamma_{i12}, \gamma_{i13})^T$, is a vector of length two. At this point, it is crucial to realize that the random effects γ_{i12} and γ_{i13} are not independent, since treatment comparisons 1 vs 2 and 1 vs 3 in the multi-arm trial i are based on the same baseline data. Therefore we need to take into account this dependency within multi-arm trials. On the other hand, we assume that inconsistency does not occur within a multi-arm trial or we say multi-arm trials are *inherently consistent*.

In order to allow for the dependency within multi-arm trials, now we assume that the heterogeneity random effect, $\boldsymbol{\gamma}_i$ follows a bivariate normal distribution as follows:

$$\boldsymbol{\gamma}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{T}_i) \tag{2.5}$$

where $\boldsymbol{T}_i$ is a two-by-two covariance matrix which is a non-diagonal matrix. This covariance matrix can have different forms. One of the simplest forms but a convenient one is proposed by Higgins and Whitehead (1996), that is a *homogeneous* covariance matrix. In the thesis we exclusively use such form of covariance matrix. One can see also Section 3.2 in Lu and Ades (2004) for a discussion about nonhomogenous covariance matrix instead of our choice. By adapting from Section 5.1 in Higgins and Whitehead (1996), we explain how such covariance matrix can be justified.

For the sake of simplicity, firstly we discuss a three-arm trial, say trial i, with treatments 1, 2 and 3. Since a multi-arm trial is inherently consistent, we have following model equations of this trial:

$$\begin{aligned}\text{logit}(\pi_{i1}) &= a_{i1} \\ \text{logit}(\pi_{i2}) &= a_{i1} + d_{12} + \gamma_{i12} \\ \text{logit}(\pi_{i3}) &= a_{i1} + d_{13} + \gamma_{i13}\end{aligned} \tag{2.6}$$

Then, if we write down the treatment comparisons 1 vs 2 and 1 vs 3 on the logit scale by following equation (2.6), we obtain:

$$\begin{aligned}\text{logit}(\pi_{i2}) - \text{logit}(\pi_{i1}) &= d_{12} + \gamma_{i12} \\ \text{logit}(\pi_{i3}) - \text{logit}(\pi_{i1}) &= d_{13} + \gamma_{i13}\end{aligned} \tag{2.7}$$

At this point consider treatment comparison 2 vs 3 on logit scale, so we get following equation:

$$\text{logit}(\pi_{i3}) - \text{logit}(\pi_{i2}) = d_{13} - d_{12} + \gamma_{i13} - \gamma_{i12} \tag{2.8}$$

The key assumption to obtain a homogeneous covariance matrix is the homogeneity of between-study variations for every treatment comparison:

$$\text{Var}(\text{logit}(\pi_{i2}) - \text{logit}(\pi_{i1})) = \text{Var}(\text{logit}(\pi_{i3}) - \text{logit}(\pi_{i1})) = \text{Var}(\text{logit}(\pi_{i3}) - \text{logit}(\pi_{i2})) \tag{2.9}$$

Since d_{13} and d_{12} are fixed variables, we can combine equation (2.7), equation (2.8) and equation (2.9) as follows:

$$\text{Var}(\gamma_{i12}) = \text{Var}(\gamma_{i13}) = \text{Var}(\gamma_{i13} - \gamma_{i12}) \tag{2.10}$$

On the other hand, since $(\gamma_{i12}, \gamma_{i13})^T$ is a bivariate random variable, we have following variance equation (see Held and Sabanés Bové, 2014a, Appendix A.3.5):

$$\text{Var}(\gamma_{i13} - \gamma_{i12}) = \text{Var}(\gamma_{i13}) + \text{Var}(\gamma_{i12}) - 2\text{Cov}(\gamma_{i13}, \gamma_{i12}) \tag{2.11}$$

Here, without loss of generality, say $\text{Var}(\gamma_{i12}) = \tau^2$ and $\text{Cov}(\gamma_{i12}, \gamma_{i13}) = \rho\tau^2$. Therefore, equation (2.10) and equation (2.11) implies that:

$$\begin{aligned}\text{Var}(\gamma_{i13} - \gamma_{i12}) &= \tau^2 \\ &= 2\tau^2 - 2\rho\tau^2\end{aligned} \tag{2.12}$$

Hence we get $\rho = 0.5$ which is the correlation between γ_{i12} and γ_{i13}. The resulting $\boldsymbol{T}_i$ is as follows:

$$\boldsymbol{T}_i = \begin{bmatrix} \tau^2 & \tau^2/2 \\ \tau^2/2 & \tau^2 \end{bmatrix}$$

If we generalize this method to a multi-arm trial i with K_i different treatments, $\boldsymbol{\gamma}_i$ is a vector of length $(K_i - 1)$ and follows a multivariate normal distribution (i.e. $\boldsymbol{\gamma}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{T}_i)$). Here, $\boldsymbol{T}_i$ is a symmetric homogeneous covariance matrix of dimension $(K_i - 1) \times (K_i - 1)$ with diagonal entries equal to τ^2 and non-diagonal entries set to $\tau^2/2$ (see also Higgins *et al.*, 2012).

With the presence of multi-arm trials, we still assume cycle-specific random effects $\xi_{jkl} \sim \mathcal{N}(0, \kappa^2)$ to account for inconsistency. In a network with also include multi-arm trials, ICDF $= \#\ (d_f) - S$ where S is the number of independent inconsistency relations in which the corresponding parameters are supported by no more than two independent source (Lu and Ades, 2006). To explain this situation better, consider a direct estimate only calculated from a multi-arm trial. Then any indirect estimate based on the other treatments in the same multi-arm trial does not form an independent cycle, since the baseline treatment is same for every comparison. For example, there is no inconsistency in a network which includes a two-arm trial with structure 1 vs 2 and a multi-arm trial with structure 1 vs 2 vs 3. The reason is that the direct estimate of comparison 1 vs 3 is only calculated from a multi-arm trial (only from one source). Therefore, we should be careful for calculation of ICDF in such situations. As a consequence, identification of all independent 3-cycles and calculating ICDF can be complicated for a network with many multi-arm trials. In such cases there is no general formula to calculate ICDF. Therefore the calculation must be done "by hand" (see section 4.2 and 4.5 in Lu and Ades, 2006).

2.4 An application (Hip dataset)

We use a network meta-analysis dataset from a recently published systematic review paper (Murad *et al.*, 2012) to illustrate the application of INLA approach by using `r-inla` package. To begin with we will describe the dataset. Then in the analysis part, firstly we will analyse the dataset with pairwise meta-analysis using different inference methods and compare their results. Finally we will conduct a network meta-analysis with INLA and compare the results obtained by the same methods used in the original paper. However we will not present systematic reviews, study selection and data extraction criterion which were reported in the primary paper as those aspects are not our focus in the thesis.

Our dataset consists of randomized controlled trials which compared treatments to prevent fragility hip fractures in individuals with or at risk of osteoporosis. Osteoporosis, characterized by low bone mineral density and deterioration of bone structure, is primarily present in postmenopausal women and is associated with an increased risk of fragility

fractures, namely hip, vertebral and other non-vertebral fractures (Deal, 1997). In the dataset, trials measure hip fragility fractures as the outcome of interest. The dataset is hereafter referred as Hip dataset. In total, there are 40 trials and 11 treatments in the Hip dataset. Table 2.1 shows the coding of treatments of interest which will be used in plotting and analysis of the network. Note that there are one three-arm trial with the structure 1 vs 9 vs 10 and two four-arm trials with the structure 1 vs 9 vs 10 vs 11.

Table 2.1: Treatments of interest in Hip dataset with their coding. These codes are used in the analysis procedure.

CODE	Treatment
1	Vitamin D+Calcium
1	Vitamin D+Calcium+Placebo
2	Teriparatide(PTH)+Vitamin D+Calcium
3	Demosumab+Vitamin D+Calcium
4	Raloxifene
4	Raloxifene+Vitamin D+Calcium
5	Zoledronate+Vitamin D+Calcium
6	Risedronate
6	Risedronate+Vitamin D+Calcium
7	Ibandronate+Calcium+Vitamin D
8	Alendronate+Calcium
8	Alendronate+Vitamin D
8	Alendronate+Vitamin D+Calcium
9	Vitamin D
9	Vitamin D+Placebo
10	Placebo
11	Calcium
11	Placebo+Calcium

2.4.1 Pairwise meta-analysis

In the original paper, it is stated that direct head-to-head comparisons (or pairwise meta-analysis) were conducted by using a random effects model with classical approach (method of moment, shortly MOM, estimator as shown in equation (1.6)). This procedure was performed using Comprehensive Meta-analysis version 2 software package (Borenstein *et al.*, 2005). To conduct a pairwise meta-analysis, we chose the treatment comparisons which have number of study is bigger than 1, namely 1 vs 8, 1 vs 6, 6 vs 10, 1 vs 4 and 1 vs 5 from given direct head-to-head comparisons in the original paper for Hip dataset.

In the thesis, in order to implement the same MOM methodology in R environment, we use the `metafor` (Viechtbauer, 2010) package, as we used in Chapter 1, but here to analyse Hip dataset. Moreover, we use a fixed effect and random effects model with INLA using summary-level data as discussed in Section 1.3.4.

By using same method which we demonstrated in Section 1.3.2, it is also possible to demonstrate an individual study using a two-by-two contingency table. In the Hip dataset, there are zero entries of a two-by-two table of some trials, namely the comparison of 1 vs 8 and 1 vs 4. The analysis by using summary level data (both MOM and INLA) requires the calculation of odds ratios which is not possible when there is a zero entry of a two-by-two table of a trial, as a result this creates a problem. Therefore to circumvent this problem, 0.5 has been added to all the cells of the two-by-two table in which there is a zero cell, as suggested by others (Cox and Snell, 1989). We used the log odds ratio as the measure of treatment effect. Table 2.2 demonstrates the point estimates both for ν and τ^2 and 95% confidence intervals for ν obtained by MOM approach. By using `metafor` package, we obtained same results for all estimates up to second digit with the results appeared in the original paper. The implementation of INLA approach was done with **meta.inla** function as we used in Chapter 1. For the model choice of INLA method, we used fixed effect model with prior distribution $\theta \sim \mathcal{N}(0, 1000)$ for comparisons 6 vs 10, 1 vs 4 and 1 vs 5. We considered a fixed effect model is sufficient since the number of studies only two for those comparisons. For comparison 1 vs 8, we conducted a random effects model with prior distributions $\nu \sim \mathcal{N}(0, 1000)$ and $\tau \sim \mathcal{U}(0, 10)$. However, for comparison 1 vs 6, we obtained substantially wider interval estimates, when we used the same prior distributions. This may occur because of the very low number of studies for that comparison which is only three. For this reason, we used more informative prior for τ, precisely $\tau \sim \mathcal{U}(0, 1)$. Table 2.2 shows the posterior mean of ν with 95% credible intervals and the posterior mean of τ^2 that are obtained by INLA.

As a conclusion, when using INLA approach as an inference method for pairwise meta-analysis, one should be careful, when the number of studies is low. Since as in our dataset, the results can be heavily affected by the prior specification especially if the number of studies is low. Although the choice of prior distributions is not our focus in the thesis, when the number of studies is low, external knowledge may be used for determination of an informative prior distributions instead of flat priors as we are mainly using in the thesis.

2.4.2 Network meta-analysis

For the statistical analysis of this network, in the reference paper, the model which was introduced by Lu and Ades (2004) was used. This model is actually very similar to the consistency model of trial-arm level approach (Section 2.3.2). In the reference paper, MCMC via WinBUGS 1.4.3 (Lunn *et al.*, 2000) was used to analyse the Hip dataset. For inconsistency, they compared the estimates from the direct comparisons and those from

Table 2.2: Results of meta-analysis of direct comparisons of Hip dataset with method of moment approach by using `metafor` (MOM) and INLA approach with summary level dataset (INLA). The number of studies for each comparison is shown (# of st.). In MOM, 2.5% and 97.5% correspond to lower bound and upper bound for a 95% confidence interval for ν, respectively. Moreover, in INLA, 2.5% and 97.5% correspond to lower bound and upper bound for a 95% credible interval for ν, respectively.

Comparison	**# of st.**	**MOM**				**INLA**			
		ν	2.5%	97.5%	τ^2	ν	2.5%	97.5%	τ^2
1 vs 8	7	-0.44	-0.87	-0.01	0	-0.47	-1.22	0.27	0.02
1 vs 6	3	-0.65	-1.46	0.17	0.27	-0.68	-1.98	0.62	0.94
6 vs 10	2	-1.78	-3.02	-0.54	NA	-1.78	-3.02	-0.53	NA
1 vs 4	2	0.11	-0.44	0.66	NA	0.11	-0.44	0.66	NA
1 vs 5	2	-0.49	-0.78	-0.2	NA	-0.49	-0.78	-0.2	NA

the indirect comparisons for the magnitude and direction of the point estimates and the extent of overlap of CI.

Here, we will fit firstly the consistency model of trial-arm level approach which we discussed in Section 2.3.2 using MCMC and then INLA. For MCMC method, we will use JAGS (Plummer *et al.*, 2003), a program for analysis of Bayesian hierarchical models using MCMC, from within R with the help of `R2jags` (Su and Yajima, 2015) R package. The same statistical model will be fitted via INLA. Then, we will implement trial-arm level approach for incorporating both heterogeneity and inconsistency (cycle-specific approach) by using INLA as described in Section 2.3.2. To fit those models via INLA, we will follow the methodology described in Sauter and Held (2015). Finally a comparison between results will be shown.

Data preparation and parameters

In order to fit a NMA model with `r-inla`, the dataset should be brought to a suitable format. This data preparation can be done by using **creatINLAdat** function from `nmainla` R package which is available for download in the online version of the Sauter and Held (2015) at the publishers web-site. The **creatINLAdat** function adds indicator variables to a data frame which define the basic contrasts, the heterogeneity random effects (`re`), the grouping vector which defines the covariance structure (`g`) (for the correlated multi-arm trials). As we discussed before, any spanning tree can be used as basic contrasts in trial-arm level approach. However due to lack of flexibility of the **creatINLAdat** function, we are allowed to use only one "non-d_{1x}" basic contrast. Luckily, in the Hip dataset, the treatment of Vitamin D+Calcium (and Vitamin D+Calcium+Placebo) is a

global baseline treatment. Note that when there is not a global baseline treatment in the network, **creatINLAdat** should be adapted. In the raw dataset, the coding of Vitamin D+Calcium was 10. Hence, we switch the coding 10 and the coding 1 in order to code the chosen global baseline treatment as treatment 1. Table 2.1 is showing the coding of treatments after this operation. In the network, the number of basic contrasts is 10, $\#\ (\mathsf{d}_b) = 10$. Also, since there are 18 available treatment comparisons, the number of functional contrasts is $18 - 10 = 8$, so $\#\ (\mathsf{d}_f) = 8$. Figure 2.4 demonstrates the plot of the corresponding network after data preparation which is called as *network graph*. Network graphs can be very useful as a data visualization technique for a NMA dataset. The corresponding R-code for arranging dataset and plotting Figure 2.4 can bee seen in Appendix A.2.1.

Although changing coding of treatments is not necessary for MCMC, we will analyse the dataset by using both method (MCMC and INLA) after these steps in order not to have any effects on the results stem from these changes. We will discuss later whether the results can be affected by the changed treatment coding or not (Section 3.5).

Consistency model

Here, we use a trial-arm level approach by assuming consistency in the network. This model, actually, can be obtained by not introducing any inconsistency random effects to the model. For the implementation in both MCMC and INLA, we used the same non-informative priors. For basic contrasts $d_b \sim \mathcal{N}(0, 1000)$, and for heterogeneity standard deviation $\tau \sim U(0, 10)$ were used.

Firstly, for the implementation of this model with JAGS, we adapted the WinBUGS code given in Jackson *et al.* (2014) to our consistency model. The corresponding JAGS code can be seen in Appendix A.2 (see Listing A.1). The R-code for the implementation of this procedure can be seen in Appendix A.2.2. A crucial component of any MCMC implementation is checking the convergence diagnostics to have an idea whether Markov chains converge to the posterior distribution. However, MCMC methods are not our focus in the thesis, so we are not going to give any details about those concepts. In our application, firstly to ensure convergence, we simulated 60 000 draws after 30 000 burn-in phase with three chains by taking those values from the original systematic review paper (Murad *et al.*, 2012). Gelman-Rubin convergence statistic (Gelman and Rubin, 1992) were stable in all instances. However, the effective sample sizes were substantially small especially for parameters of d_{12} and τ (this situation may stem from high autocorrelation). To overcome this problem, we increased the number of simulations to 600 000 draws after 300 000 burn-in by using 3 as thinning parameter and again with three chains. Then, we checked convergence diagnostics by using Gelman-Rubin convergence statistic, visual inspection of autocorrelation histogram plots and traceplots. It seems that the number of simulations are sufficient. Table 2.3 shows the posterior median and 95 % credible interval (Cr. I) obtained by MCMC for basic contrasts and heterogeneity standard deviation.

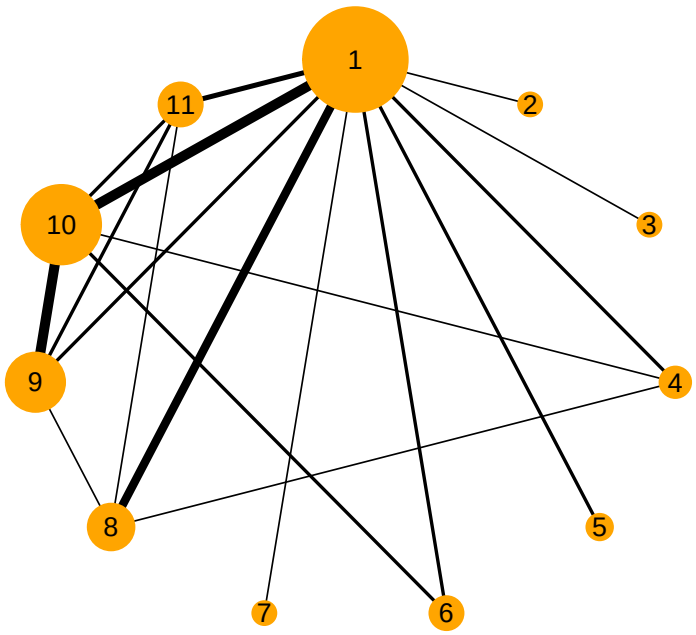

Figure 2.4: Network of trials evaluating hip fractures (Hip dataset). The size of the circle is proportional to the number of participants to that treatment. Width of lines is proportional to the number of trials for that comparison.

Moreover, these results are consistent with the results appeared in the original paper.

For INLA, we must rearrange the dataset to get a trial-arm level (one-arm-per-row) by using the **mtc.data.studyrow** function from the `gemtc` R package. Then by using the **creatINLAdat** function, the dataset can be brought to a suitable format to analyse with `r-inla`. For incorporating multi-arm trials, we want to have a homogeneous covariance matrix, $\boldsymbol{T}_i$, to correlate the random effects for all arms of the same trial. As explained in Section 3.2 of Sauter and Held (2015), the homogeneous covariance matrix which we need is implemented in `r-inla` under the name `model ="exchangeable"`. Also, we need to define a grouping vector g_i to determine the structure of $\boldsymbol{T}_i$. The first entry of g_i is `NA` as no random effect is present in the baseline treatment. The remaining entries are numbered as $1, 2, \ldots, K_i - 1$ to represent heterogeneity random effects $\gamma_{ij2}, \gamma_{ij3}, \ldots, \gamma_{ij(K_i-1)}$ (Sauter and Held, 2015). Also, this grouping vector (`g`) is obtained by **creatINLAdat** function. The corresponding R-code can be seen in Appendix A.2.3.

The results are shown in Table 2.3. The comparison of the results estimated by MCMC

and INLA can be seen in Figure 2.5. It can be said that MCMC and INLA results show very good agreement. For both MCMC and INLA methods, functional contrasts can be calculated by using consistency relations as explained in Section 2.2.3.

Table 2.3: Results of the consistency model implemented by MCMC and INLA for Hip dataset. The last line shows the estimates for random effects standard deviation of heterogeneity.

	MCMC			**INLA**		
	Median	2.5%	97.5%	Median	2.5%	97.5%
$d_{1,2}$	-0.709	-2.172	0.791	-0.703	-2.121	0.714
$d_{1,3}$	-0.517	-1.082	0.028	-0.506	-1.079	0.049
$d_{1,4}$	0.068	-0.263	0.419	0.067	-0.283	0.431
$d_{1,5}$	-0.487	-0.833	-0.141	-0.484	-0.844	-0.126
$d_{1,6}$	-0.522	-0.926	-0.234	-0.552	-0.936	-0.238
$d_{1,7}$	-0.529	-1.380	0.342	-0.539	-1.385	0.322
$d_{1,8}$	-0.600	-1.019	-0.195	-0.598	-1.024	-0.190
$d_{1,9}$	0.327	0.100	0.562	0.330	0.090	0.574
$d_{1,10}$	0.207	0.040	0.387	0.210	0.030	0.399
$d_{1,11}$	0.344	0.031	0.667	0.349	0.021	0.679
τ	0.085	0.004	0.279	0.087	0.007	0.276

Inconsistency model

In order to account for inconsistency in the network of the Hip dataset, we will implement the inconsistency model using trial-arm level approach (cycle-specific approach) with INLA. To achieve this, firstly we must determine ICDF by hand and its corresponding inconsistency relations. If we examine the network, we can realize that two of the functional parameters $d_{9,11}$ and $d_{10,11}$ are only estimated by four-arm trials. This gives ICDF $= \#\,(\mathsf{d}_f) - \mathrm{S} = 8 - 2 = 6$. The corresponding inconsistency relations are as follows:

$$
\begin{aligned}
d_{4,8} &= d_{1,8} - d_{1,4} + \xi_{1,4,8} \\
d_{4,10} &= d_{1,10} - d_{1,4} + \xi_{1,4,10} \\
d_{6,10} &= d_{1,10} - d_{1,6} + \xi_{1,6,10} \\
d_{8,9} &= d_{1,9} - d_{1,8} + \xi_{1,8,9} \\
d_{8,11} &= d_{1,11} - d_{1,8} + \xi_{1,8,11} \\
d_{9,10} &= d_{1,10} - d_{1,9} + \xi_{1,9,10}
\end{aligned}
\tag{2.13}
$$

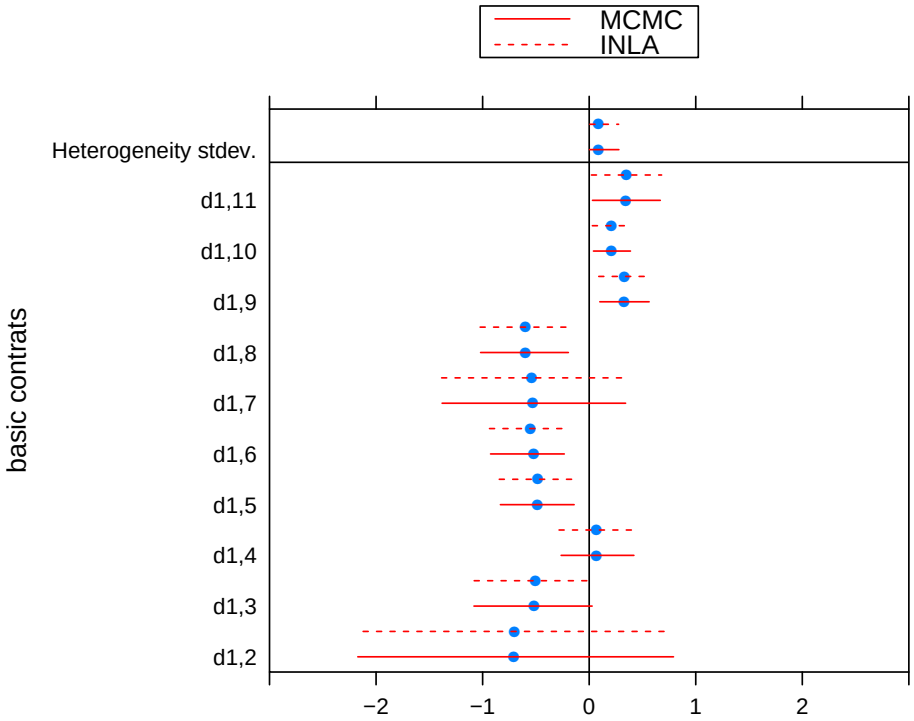

Figure 2.5: Graphical representation of the results of the implemented consistency model for Hip dataset by using MCMC and INLA. The estimates for random effects standard deviation of heterogeneity is shown at the top (Heterogeneity stdev.). The points correspond to posterior medians of each parameter. The corresponding lines refer to 95 % equi-tailed credible intervals of each parameter.

In order to implement this method in `r-inla`, the inconsistency random effects should be given introduced in advance. For this purpose, we add an indicator variable, `w`, to the data frame which define the above cycle-specific inconsistency random effects. Moreover, so as to estimate functional contrasts, we need to define linear combinations for each of them. To achieve this, we used the function **inla.make.lincombs** from `r-inla`. We used vague prior for inconsistency standard deviation $\kappa \sim U(0, 10)$. The corresponding R-code can be seen in Appendix A.3.1. Table 2.4 and Table 2.5 show the posterior median and 95 % credible interval (Cr. I) obtained by this cycle-specific approach for basic contrasts and functional contrasts, respectively. Also estimates for heterogeneity and inconsistency random effects standard deviations are shown. For inconsistency standard deviation, we obtained posterior median estimate of 0.306 with 95 % credible interval from 0.016 to 1.405. Therefore, there is no evidence which suggests a severe inconsistency in the network. Furthermore, when we compare Table 2.3 and Table 2.4, we can see that introducing inconsistency random effects does not have a large impact on the estimates of basic contrasts.

Estimates of the inconsistency random effect parameters may also shed light on the presence of the inconsistency within this network. Figure 2.6 displays posterior median estimates with corresponding 95 % credible interval for inconsistency random effects. We can conclude that there is no evidence advocating a severe inconsistency in the network when we use a cycle-specific approach. This conclusion is also same as in the conclusion of the reference paper in which they stated that there is no "significant" inconsistency in the network.

Table 2.4: The results of the inconsistency model by using cycle-specific approach with INLA for Hip dataset. The last lines show the estimates for random effects standard deviation of heterogeneity and inconsistency, respectively.

	Median	2.5%	97.5%
$d_{1,2}$	-0.703	-2.126	0.718
$d_{1,3}$	-0.507	-1.091	0.060
$d_{1,4}$	0.103	-0.388	0.645
$d_{1,5}$	-0.484	-0.853	-0.115
$d_{1,6}$	-0.517	-0.942	-0.187
$d_{1,7}$	-0.539	-1.392	0.330
$d_{1,8}$	-0.582	-1.029	-0.152
$d_{1,9}$	0.228	-0.192	0.584
$d_{1,10}$	0.204	0.008	0.411
$d_{1,11}$	0.321	-0.020	0.667
τ	0.087	0.006	0.278
κ	0.306	0.016	1.405

Table 2.5: The estimates of functional contrasts using cycle-specific approach with INLA.

	Median	2.5%	97.5%
$d_{4,8}$	-0.693	-1.926	0.472
$d_{4,10}$	0.157	-0.233	0.544
$d_{6,10}$	1.076	0.435	2.224
$d_{8,9}$	1.000	0.284	1.936
$d_{8,11}$	1.001	0.005	2.239
$d_{9,10}$	-0.143	-0.349	0.063

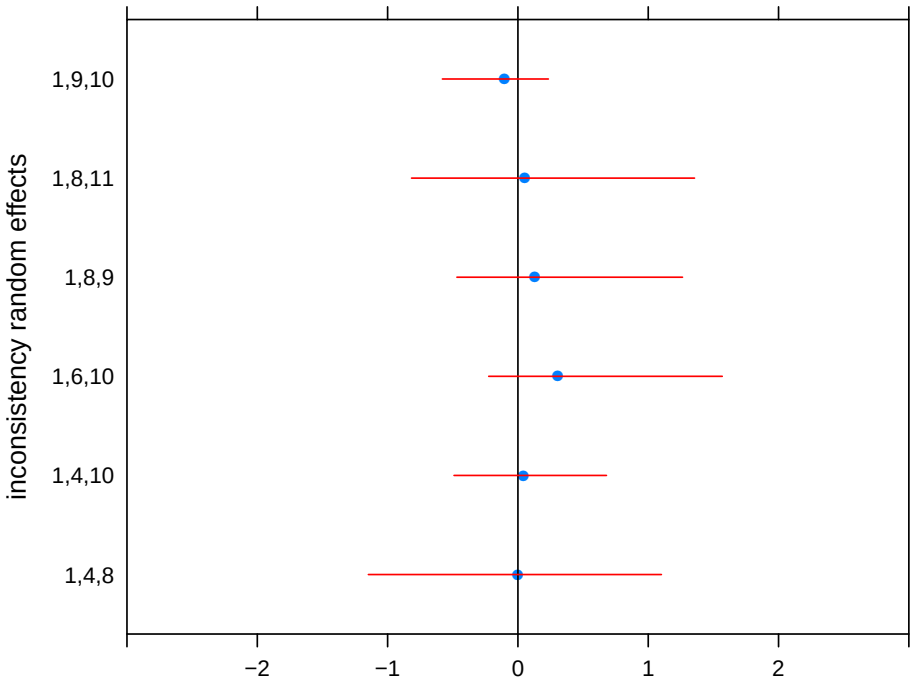

Figure 2.6: Estimates of the inconsistency random effects parameters using cycle-specific approach with INLA.

R version and packages used to generate this chapter:

R version: R version 3.2.3 (2015-12-10)

Base packages: splines, stats, graphics, grDevices, utils, datasets, methods, base

Other packages: lattice, metafor, pcnetmeta, nmainla, R2jags, rjags, gemtc, coda, INLA, Matrix, sp, igraph, xtable

Versions of other packages (respectively): 0.20.33, 1.9.8, 2.3, 1.1, 0.5.7, 4.5, 0.7.1, 0.18.1, 0.0.1443538834, 1.2.3, 1.2.2, 1.0.1, 1.8.2

This document was generated on February 23, 2016 at 13:24.

Chapter 3

Design-by-treatment interaction model

3.1 Introduction

As opposed to pairwise meta-analysis, network meta-analysis offers the advantage of being able to compare any treatment included in the network, including those that have not been compared directly. However, there are some issues concerning network meta-analysis. For example, the usage of more complicated statistical models to incorporate indirect estimates of treatment comparisons presets problems. According to Song *et al.* (2009), there is a variety of problems associated with network meta-analysis, including "unclear understanding of underlying assumptions", "use of flawed or inappropriate methods" and "inadequate comparison and inappropriate combination of direct and indirect evidence".

From a statistical perspective, one of the biggest challenges facing network meta-analysis is inconsistency (or incoherence). As we discussed in Chapter 2, this occurs when there is a discrepancy between direct and indirect estimates. We described two different statistical models, summary level approach or *Lumley model* (Lumley, 2002), and trial-arm level approach or *Lu-Ades model* (Lu and Ades, 2006), to incorporate inconsistency in the network. Sauter and Held (2015) showed that INLA can be used as an alternative inference method both for Lumley or Lu-Ades models which we discussed in Chapter 2. Moreover, another method for assessing the inconsistency in the network is *node-splitting* where inferences are split depending on whether the information comes from studies that provide direct estimates or indirect estimates about a particular relative treatment effect (Dias *et al.*, 2010). The INLA implementation of the node-splitting is possible and introduced by Sauter and Held (2015). However, in the thesis we want to discuss another method for assessing inconsistency instead of node-splitting.

Very recently a new statistical model, called the *design-by-treatment interaction model*, is introduced by Higgins *et al.* (2012). In their companion paper (White *et al.*, 2012), they applied this new model with fixed inconsistency parameters in a frequentist approach and

in a Bayesian approach via MCMC. After that, Jackson *et al.* (2014) introduced design-by-treatment interaction model with random inconsistency parameters via only MCMC.

In this chapter we firstly describe the design inconsistency and corresponding statistical model for NMA, then by using both MCMC and INLA as inference methods we will apply design-by-treatment interaction model with random inconsistency parameters and compare the results.

3.2 Cycle inconsistency and design inconsistency

The following explanations are mostly adapted from Higgins *et al.* (2012) and Jackson *et al.* (2014). In this section, we discuss the term *design*, *design inconsistency* and its relation to *cycle inconsistency*, also we used a graphical representation, Figure 3.1, of different examples to make explanations easier. However, note that this graphical representation is different than a network graph which we used in Chapter 2.

3.2.1 Consistency

In order to make it easier to comprehend, we illustrate different concepts with different scenarios. Each scenario include three treatments namely A, B and C. We start with a simple scenario with a network only include two-arm trials with the structure A vs B, A vs C and B vs C under the consistency assumption. Therefore, the consistency in such a network can be expressed with a consistency relation:

$$d_{AB} = d_{AC} - d_{BC} \tag{3.1}$$

where d_{jk} parameters represent mean treatment effect across all trials of comparison j vs k as we discussed in Section 2.2.3. Figure 3.1a shows this scenario (a network with only two-arm trials in which the consistency assumption holds). We draw all edges by using the same solid line style in Figure 3.1a to show that there is no inconsistency in the cycle.

3.2.2 Cycle inconsistency

The consistency may not hold if there is a substantial difference between estimates of treatment comparisons so that treatment effects do not "add up" around the cycle in the figure. Figure 3.1b demonstrates such cycle inconsistency situation when the network includes only two-arm trials. The single dashed line in Figure 3.1b is used to indicate that there is inconsistency in the network. However, the place of the single dashed line in Figure 3.1b is arbitrary, since the cycle inconsistency is a property of a cycle rather than a pairwise treatment comparison. Therefore, as an alternative scenario to Figure 3.1b one can also demonstrate the same cycle inconsistency situation using a different graph. For

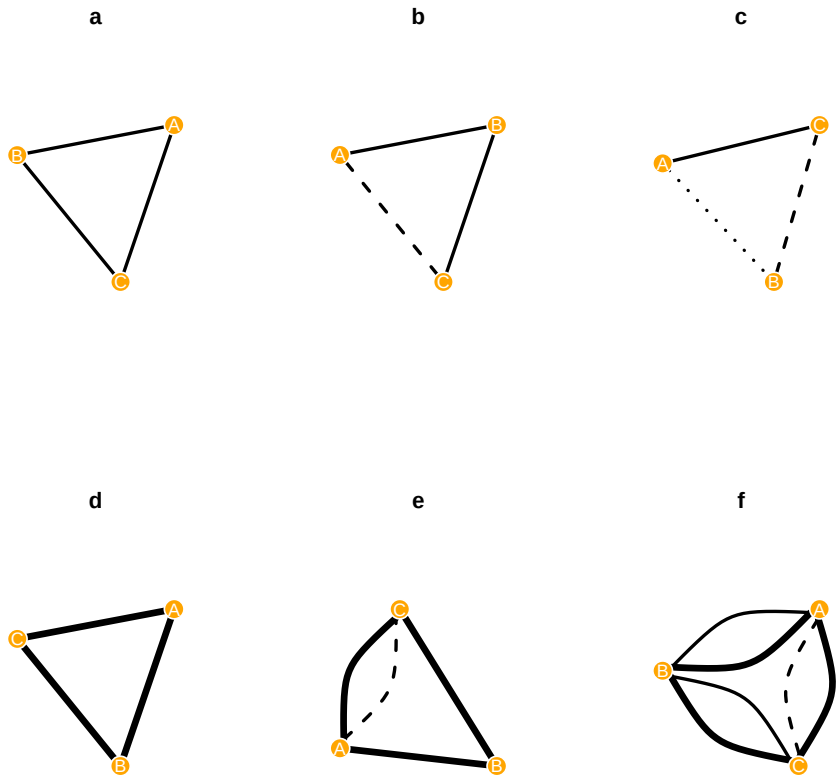

Figure 3.1: Graphical representation of consistency, cycle inconsistency and design inconsistency. (**a**) consistency: the consistency assumption holds in the network of two-arm trials only. (**b**) cycle inconsistency: there is discrepancy between direct estimate coming from A vs C (dashed line) and indirect estimate coming from A vs B and B vs C (solid lines). (**c**) cycle inconsistency: alternative scenario, indistinguishable from (**b**) without additional evidence. (**d**) consistency: three-arm trial is inherently consistent. (**e**) design inconsistency: there is discrepancy between estimate coming from three-arm trial and from the two-arm trials. (**f**) design inconsistency and cycle inconsistency: cycle inconsistency within the cycle of two-arm trials, whereas the three-arm trial conflicts with at least one two-arm trial which is design inconsistency.

instance, Figure 3.1c displays the same network as in Figure 3.1b, but now three edges are shown in different line styles in order to indicate different possible sources of inconsistency that are associated with each edge in the cycle. However, investigation of this kind of alternative scenario is not our focus, since it cannot be tested statistically.

3.2.3 Multi-arm trials

The presence of multi-arm trials in the network complicates the cycle inconsistency approach as we emphasized in Chapter 2. This is because the calculation of the ICDF becomes hard. Within a multi-arm trial, cycle inconsistency cannot occur by definition. Figure 3.1d shows a network which includes only three-arm trials. In Figure 3.1d, three edges are drawn as thick lines to indicate three-arm trials. Also, the same line style is used for all edges to display that there is no inconsistency in the network. Therefore, within the framework of cycle inconsistency, a network may be consistent either structurally, since all studies include all treatments (as in Figure 3.1d), or by observation, when consistency assumption holds (as in Figure 3.1a), or by a combination of this two.

3.2.4 Design inconsistency

By the *design* of a trial, we mean the set of treatments included in a trial. *Design inconsistency* refers to differences in treatment effects between trials involving different designs. In allowing this difference, we implicitly assume that different designs may serve as a proxy for some important effect modifiers. In Figure 3.1e, a design inconsistency is shown. The three thick lines correspond to a three-arm trial with the design of A vs B vs C. By using dashed line of AC edge, we imply that there is a design inconsistency between AC treatment comparison in the two-arm trial (dashed line) and the AC treatment comparison in three-arm trial (thick line).

From a different perspective, design inconsistency can be considered as a special case of heterogeneity, since the study designs correspond to a study-level covariate which has an association with treatment effect as in a pairwise meta-regression (see Section 1.4). The design inconsistency approach is very helpful for solving the problems occurring with the presence of multi-arm trials in the network. For a network with only two-arm trials, however, the concept of design inconsistency provides no added insights compared with cycle inconsistency.

In Figure 3.1f, an example in which a cycle inconsistency implies design inconsistency is shown. Three thick lines in this figure correspond to a three-arm trial with design A vs B vs C. All other three edges in this graph correspond to two-arm trials with designs A vs B, A vs C and B vs C. From cycle-inconsistency perspective, the dashed line correspond to a cycle inconsistency in the cycle (ABC) formed by two-arm trials. On the other hand, since the three-arm trial is inherently consistent, there is design inconsistency between treatment effect estimated by the three-arm trial and at least one of the two-arm trials in this scenario. However, when the network has multiple multi-arm trials and two-arm trials, the distinction between cycle inconsistency and design inconsistency becomes complicated as Higgins *et al.* (2012) pointed out. The *full design-by-treatment interaction model*, which we will introduce now, encompasses both cycle inconsistency and design inconsistency.

3.3 The Design-by-treatment interaction model

It is possible to have the design-by-treatment interaction model with random effects formulation either using a summary level approach (contrasts-based) or a trial-arm level approach (arm-based). The model using a summary level approach was widely described in Section 2 of Jackson *et al.* (2014). That model can be considered as a multivariate extension of the Lumley model, in which it is assumed that true treatment effects follow a multivariate normal distribution. However, in the thesis we prefer to discuss exclusively the trial-arm level approach. The consistency model of the Lu-Ades model and the consistency model of the design-by-interaction model with trial-arm level is actually same, since two models differ in defining inconsistency parameters in the network. Therefore this model can be considered as an extension of the consistency model of the Lu-Ades model (Section 2.3.2). Actually, the design-by-treatment interaction model contains additional inconsistency random effects compared to the Lu-Ades model.

The model can be written as follows: Each trial $s = 1, 2, \ldots, \bar{S}$ of each design $D = 1, 2, \ldots, \bar{D}$ has treatment arms $t_1(s)^D, t_2(s)^D, \ldots, t_{K_s}(s)^D$ with at least $K_s \geq 2$ treatment arms. The first treatment $j = t_1(s)^D$ is chosen as baseline treatment and compared with the remaining treatments $k = t_2(s)^D, t_3(s)^D, \ldots, t_{K_s}(s)^D$. For each sth trial of design D and baseline treatment j the number of events y_{sj}^D and number of patients n_{sj}^D are observed. Correspondingly, also for the remaining treatments y_{sk}^D and n_{sk}^D are observed. The number of events is (conditionally) independent for each trial-arm and follows a binomial distribution, i.e. $y_{sj}^D \sim \text{Bin}(n_{sj}^D, \pi_{sj}^D)$ and $y_{sk}^D \sim \text{Bin}(\pi_{sk}^D, n_{sk}^D)$. The relative treatment effect d_{jk} of baseline treatment j vs. treatment k can now be modelled with logistic regression as:

$$\begin{aligned} \text{logit}(\pi_{sj}^D) &= a_{sj}^D \\ \text{logit}(\pi_{sk}^D) &= a_{sj}^D + d_{jk} + \gamma_{sjk}^D + \omega_{jk}^D \end{aligned} \tag{3.2}$$

The relative treatment effect a_{sj}^D of baseline treatment j in trial s of design D is a nuisance parameter and the main interest is in d_{jk}. As we had in the Lu-Ades model (Section 2.3.2), here also we assume a homogeneous covariance matrix for heterogeneity as follows:

$$\boldsymbol{\gamma}_s^D \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{\Sigma}_\gamma) \tag{3.3}$$

where $\boldsymbol{\Sigma}_\gamma$ denotes a square matrix where the diagonal entries are all τ^2 and all other entries are $\tau^2/2$, $\boldsymbol{\gamma}_s^D = (\gamma_{s12}^D, \gamma_{s13}^D, \ldots, \gamma_{s1c}^D)^T$. $\mathcal{N}_c$ denotes a multivariate normal distribution in c dimensions where c is the number of treatment arms of sth trial of design D. The extent of heterogeneity is described by τ^2. Note that here we prefer to use $\boldsymbol{\Sigma}_\gamma$ instead of $\boldsymbol{T}_i$ which was used in Section 2.3.2.

The key difference of this model and the Lu-Ades model is characterizing the inconsistency parameters. Jackson *et al.* (2014) proposed a random effects formulation of inconsistency parameters in a similar way to heterogeneity parameters as follows:

$$\boldsymbol{\omega}^D \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{\Sigma}_\omega) \tag{3.4}$$

where $\boldsymbol{\Sigma}_\omega$ denotes a square matrix where the diagonal entries are all κ^2 and all other entries are $\kappa^2/2$, $\boldsymbol{\omega}^D = (\omega_{12}^D, \omega_{13}^D, \ldots, \omega_{1c}^D)^T$. Here c is the number of treatment arms of design D. Note that by definition of a design, all trials with the same design has the same number of treatment arms. In this model, $\boldsymbol{\Sigma}_\omega$ is also a homogeneous covariance matrix. Therefore, we implicitly assume that the inconsistency variance κ^2 across designs is the same for all treatment comparisons. As we had in cycle-specific inconsistency approach (Lu-Ades model), κ^2 quantifies the extent of the inconsistency in the network as whole, and specific inconsistency parameters describe where particular inconsistencies arise. If $\kappa^2 = 0$, then there is no inconsistency, and the model becomes a consistency model. Moreover, this model can be extended to incorporate more complicated forms of $\boldsymbol{\Sigma}_\gamma$ and $\boldsymbol{\Sigma}_\omega$ in situations where it makes more sense.

The model which we just described (Jackson *et al.*, 2014) is treating inconsistency parameters ω_{jk}^D as random effects. By this way, inconsistency across evidence sources can be conceptualized as an additional variation, in the same way as heterogeneity across studies. A case for treating inconsistency parameters as fixed effects is made by Higgins *et al.* (2012). There are advantages and disadvantages of using fixed effects model formulation. One important practical advantage of using fixed effects formulation is that the model can be fitted using multivariate meta-regression in a frequentist way (White *et al.*, 2012). On the other hand, maybe the most important advantage of using random effects formulation is that we can estimate average treatment effects across all designs. However, with the usage of fixed effects for inconsistency parameters, instead of average treatment effects across all designs, the design-specific treatment effects are estimated. Hence interpretation becomes a challenging task with the usage of fixed effect formulation. Hereafter, we use the *Jackson model* to refer the design-by-treatment interaction model with random effects model formulation.

3.4 An application (Smoking dataset)

In this section, we present a NMA using the Jackson model with trial-arm level. Also the consistency model will be illustrated. We assume the same prior distributions as we had in Chapter 2. Specifically, for all basic contrasts $d_b \sim \mathcal{N}(0, 1000)$, and for heterogeneity and inconsistency standard deviations $\tau \sim U(0, 10)$ were assumed. We will apply the models to a dataset which includes 24 trials investigating interventions to aid smoking cessation, which we call the *Smoking dataset*. The smoking dataset is originally investigated by

Hasselblad (1998) but also by Dias *et al.* (2010), Higgins *et al.* (2012) and Sauter and Held (2015) among others. It must be noted that we obtained the dataset from the `nmainla` R-package and its exact source is Dias *et al.* (2010).

The Smoking dataset counts the number of individuals who successfully quit smoking after 6 to 12 months. The dataset describes a fully connected network comparing effects of four different interventions (1: self-help, 2: individual counseling, 3: group counseling and 4: no contact) and reports the number of successes and the number of participants in 24 trials. There are two three-arm trials, one for treatments 1, 3 and 4 and one for treatments 2, 3 and 4. There are eight different designs in the network which is shown in Table 3.1. Figure 3.2 demonstrates corresponding network graph. We choose intervention 1 as global baseline treatment, therefore $\mathbf{d}_b = (d_{12}, d_{13}, d_{14})^T$.

Firstly we analyze the Smoking dataset by fitting the consistency model with MCMC and INLA (as we did in Chapter 2 for the Hip dataset, see Section 2.4.2). By taking the values from Sauter and Held (2015), 20 000 iterations with an additional burn-in of 30 000 iterations are used for MCMC analysis. Figure 3.3 displays the posterior medians and the 95% equitailed credible intervals obtained by MCMC and INLA for basic contrasts and the heterogeneity standard deviation. The results show very good agreement. Moreover, the median and the 95%-CrI for τ in Figure 3.3 show that there is substantial heterogeneity present in the network.

Table 3.1: The coding of the interventions and designs in the Smoking dataset.

Design	**Intervention**			
	No contact	Self help	Individual counseling	Group counseling
1	1		3	4
2		2	3	4
3	1		3	
4	1	2		
5	1			4
6		2	3	
7		2		4
8			3	4

3.4.1 The Jackson model

The implementation of the Jackson model (the design-by-treatment interaction model with random inconsistency parameters) using trial-arm level via MCMC is achieved by JAGS from within R. The corresponding WinBUGS code were taken from the supplementary material of Jackson *et al.* (2014) and can be seen in Appendix A.3 (Listing A.2).

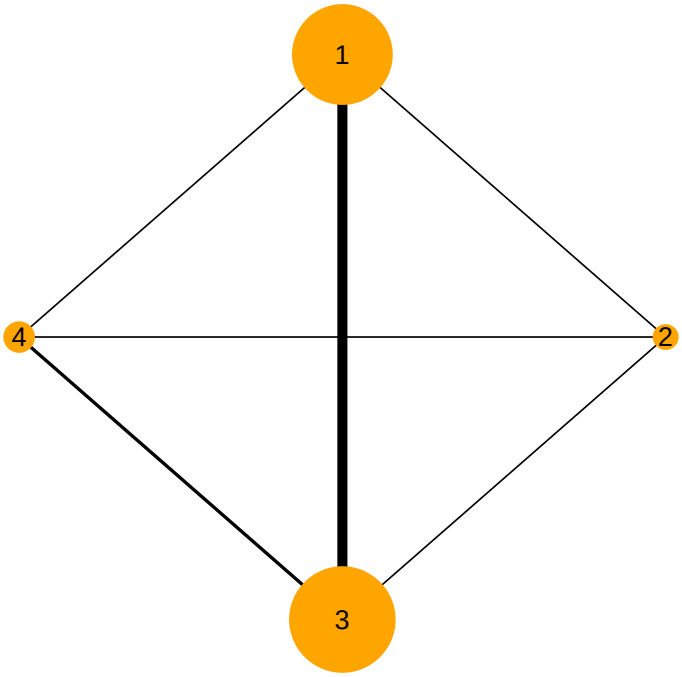

Figure 3.2: Network graph of the Smoking dataset. The size of the circle is proportional to the number of participants to that intervention. Width of lines is proportional to the number of trials for that comparison.

To get the posterior distribution of each parameter, 300 000 iterations after burn-in of 100 000 iterations were used. The R-code for the implementation of this procedure can be seen in Appendix A.4.1. Convergence checked by running three different chains by using five as thinning parameter. Gelman-Rubin convergence statistic were stable in all parameters. Also, convergence and autocorrelation were checked by visual inspection of traceplots and autocorrelation histogram plots which seems reasonable for all parameters. The histograms in Figures 3.4 shows the marginal posterior densities for all basic contrasts, heterogeneity variance and inconsistency variance that are estimated by MCMC.

Now, we explain how the Jackson model is fitted by INLA. In general, the difference between the consistency model and the Jackson model is the assumption of consistency. In the consistency model, we assume $\kappa^2 = 0$. In order to implement the latter model, the inconsistency random effects should be taken into account by `r-inla`. As we showed in equation (3.4), the assumption of inconsistency random effects is very similar to heterogeneity random effects. To implement the Jackson model, firstly we need to correlate ran-

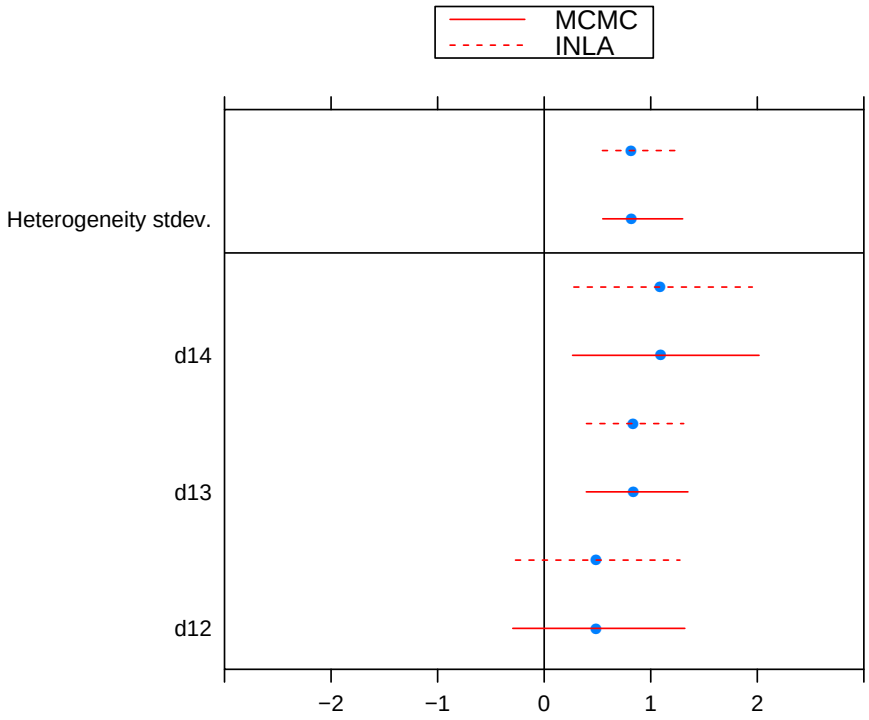

Figure 3.3: Graphical representation of the results of the implemented consistency model for the Smoking dataset with MCMC and INLA. The estimates for random effects standard deviation of heterogeneity is shown at the top (Heterogeneity stdev.).

dom effects in multi-arm trials and we achieved this by using `model = "exchangeable"` under `r-inla` as we did for heterogeneity. Then to identify the inconsistency random effects in each trial, we use the same grouping vector g_i which is needed to determine the structure of $\mathbf{\Sigma}_\omega$. The difference between inconsistency and heterogeneity random effects is that for inconsistency, the variance across designs is assumed to be same rather than a common variance across each study which is the case for heterogeneity. Therefore, we need to define an indicator variable which should be design-specific. By inspiring from the indicator variable for heterogeneity random effects (`re`), we created a new indicator variable for design-specific random effects (`des`). The variable `des` is obtained from the coding of the designs in the network. For Smoking dataset this coding of designs is shown in Table 3.1. Then, the inference is made by including inconsistency random effects to the **inla** function as we did for heterogeneity random effects. The corresponding R-code can be seen in Appendix A.4.2.

Posterior medians and 95 % credible intervals of each parameter which are obtained from the fitted Jackson model (and the consistency model) using both MCMC and INLA are shown in Table 3.2. Moreover, the straight lines in Figure 3.4 displays the marginal densities for all basic contrasts, heterogeneity variance and inconsistency variance ob-

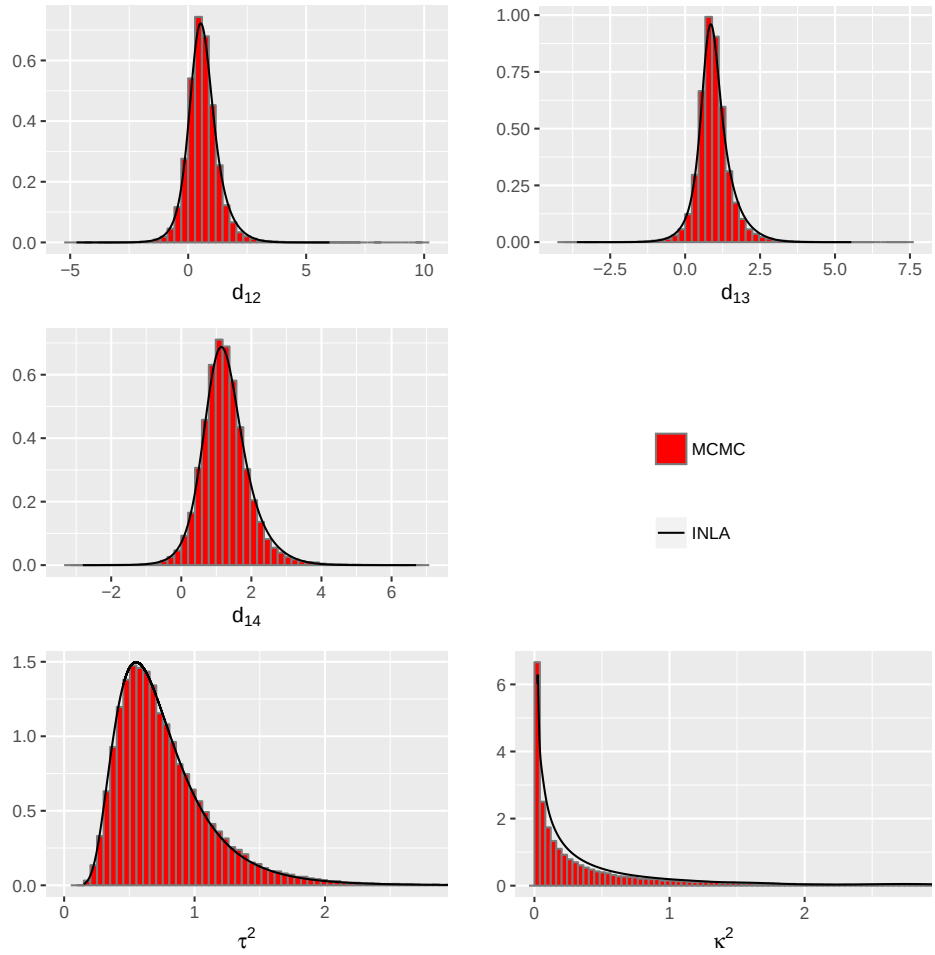

Figure 3.4: Marginal posterior density estimates of all basic contrasts, the heterogeneity and inconsistency variances by MCMC (histogram) and by INLA (straight line) obtained from the fitted Jackson model for the Smoking dataset.

tained from the INLA implementation of the Jackson model. As Figure 3.4 suggests both methods MCMC and INLA produce very similar results. Also these results are consistent with the results appeared in Table III in Jackson *et al.* (2014). Figure 3.4 displays that there is no evidence suggesting a large inconsistency in the network. Estimates of individual inconsistency parameters ω_{jk}^{D} also can be helpful to investigate the presence and even source of the inconsistency in the network. Table 3.3 shows the estimates of those parameters obtained by MCMC and INLA. We can conclude that both inference techniques give similar results.

3.5 The Lu-Ades vs the Jackson model

After we showed how INLA can be used to fit the Jackson model, in this section we present comparison between the Lu-Ades models and the Jackson models derived from different treatment orderings for Smoking dataset. This comparison is inspired from the discussion

Table 3.2: Two sets of results for the Smoking dataset, consistency model assumes consistency in the network, Jackson model allowing inconsistency in the network. Posterior median and 95 % credible interval estimates are shown for all parameters.

	Consistency model			**Jackson model**		
	Median	2.5%	97.5%	Median	2.5%	97.5%
MCMC						
$d_{1,2}$	0.487	-0.295	1.320	0.563	-0.565	2.005
$d_{1,3}$	0.838	0.395	1.349	0.901	-0.021	2.168
$d_{1,4}$	1.095	0.266	2.016	1.194	0.075	2.702
τ	0.819	0.550	1.300	0.829	0.549	1.311
κ				0.401	0.018	1.906
INLA						
$d_{1,2}$	0.487	-0.269	1.274	0.586	-0.634	2.126
$d_{1,3}$	0.833	0.397	1.309	0.916	-0.114	2.297
$d_{1,4}$	1.088	0.278	1.953	1.216	0.031	2.793
τ	0.814	0.547	1.266	0.824	0.545	1.302
κ				0.490	0.161	1.552

Table 3.3: Estimated inconsistency parameters obtained from the fitted Jackson model for the Smoking dataset.

Design	**Parameter**	**MCMC**			**INLA**		
		Median	2.5%	97.5%	Median	2.5%	97.5%
1	ω^1_{13}	0.01	-1.19	1.19	0.02	-1.33	1.33
	ω^1_{14}	-0.1	-2.06	0.65	-0.17	-2.22	0.72
2	ω^2_{23}	-0.02	-1.37	1.04	-0.04	-1.54	1.16
	ω^2_{24}	-0.03	-1.46	0.99	-0.06	-1.63	1.09
3	ω^3_{13}	-0.03	-1.34	0.84	-0.06	-1.47	0.94
4	ω^4_{12}	-0.04	-1.53	0.87	-0.08	-1.69	0.98
5	ω^5_{14}	0.16	-0.53	2.69	0.26	-0.58	2.75
6	ω^6_{23}	-0.04	-1.48	0.97	-0.07	-1.64	1.07
7	ω^7_{24}	0.03	-1.01	1.45	0.06	-1.13	1.59
8	ω^8_{34}	-0.01	-1.23	1.02	-0.02	-1.35	1.13

in Section 4 of Higgins *et al.* (2012). However, here we implement the Jackson model with INLA whereas in Higgins *et al.* (2012), the same model with fixed inconsistency parameters with frequentist approach were used.

The dataset which we analyzed in Section 3.4 had the treatment ordering 1234 (say default ordering). In total, there are 24 possible treatment orderings since there are 4 different interventions in the network of Smoking dataset. When we fit the consistency model for all 24 possible treatment orderings, the estimates of heterogeneity standard deviation appears to be same. That shows that the results do not depend on the treatment ordering for Smoking dataset (that means also the results do not depend on the chosen basic contrasts). On the other hand, the Jackson model also gives the same results of heterogeneity and inconsistency standard deviation regardless of the treatment ordering. Also we fit various Lu-Ades models using INLA (as we explained in Section 2.4.2 for the Hip dataset). The main distinction of a Lu-Ades model and a Jackson model for Smoking dataset is that the former has three inconsistency random effects parameters (since there are three independent 3-cycles in the network) whereas the latter has 10 as we showed in Table 3.3 and naturally consistency model has zero.

The results of the fitted consistency model, the Jackson model and several Lu-Ades models are shown in Table 3.4. It turns out that Lu-Ades models' results substantially depend on the treatment ordering (also the chosen basic contrasts). Especially the estimate of inconsistency standard deviation vary considerably amongst different Lu Ades models. Actually not every fitted Lu-Ades model gives rise to different results, hence it seems that some of the models are identical as is shown by the same row in Table 3.4. In fact none of the Lu-ades models provides convincing evidence of inconsistency in the network, since the lower bound of their interval estimates are very close to zero. This conclusion is also the same for the Jackson model. However, the most crucial point is that the Jackson model takes into account all possible sources of inconsistency in this network. On the other hand, there are several distinct Lu-Ades models for this dataset. Using the Jackson model, however, for a network with only two arm trials may lead to overparametrisation, hence in such cases the Lu-Ades model can be preferred. However, with the presence of multi-arm trials in the network (which is very often the case) one should avoid the Lu-Ades model because the results depend on the treatment ordering.

Very recently, Jackson *et al.* (2015) showed that the Lu-Ades model is a restricted version of the design-by-treatment interaction model and the design-by-treatment interaction model is a unifying framework for modelling cycle inconsistency in network meta-analysis. They made this conclusion by proving following statement: "The only model that contains all the Lu-Ades models with all different treatment orderings is the design-by-treatment interaction model". However we are skipping more details about this discussion since it is beyond the scope of the thesis and referring to Jackson *et al.* (2015) for interested readers.

Since the design-by-treatment interaction model should be preferred with the presence of multi-arm trials, in the next section we will fit this kind of model to the Hip dataset which we have already analyzed using the Lu-Ades model in Section 2.4.2.

Table 3.4: Estimated heterogeneity standard deviation (τ) and inconsistency standard deviation (κ) for the consistency model, the Jackson model and various Lu-ades models for Smoking dataset with INLA. Also the number of inconsistency parameters for each model is shown (# of inc. parameters). Model '1234' is the default ordering which was introduced in Table 3.1. Other models has different treatment orderings. For this particular dataset, the models grouped in rows turn out to be identical models.

	# of inc. parameters	κ			τ		
		Median	2.5%	97.5%	Median	2.5%	97.5%
Consistency model	0	0.00	0.00	0.00	0.81	0.55	1.27
Jackson model	10	0.49	0.16	1.55	0.82	0.55	1.30
Lu-ades models							
1234, 1243	3	0.54	0.03	3.56	0.84	0.56	1.32
1324, 1423	3	0.62	0.04	3.92	0.83	0.56	1.31
1342, 1432	3	0.57	0.03	3.82	0.84	0.56	1.32
2314, 3214	3	2.01	0.12	5.44	0.79	0.53	1.24
3412, 4213	3	2.04	0.12	5.56	0.79	0.53	1.24
2143, 2431	3	0.53	0.03	3.59	0.84	0.56	1.32
2341, 2413, 3241 3421, 4231	3	0.60	0.03	4.22	0.84	0.56	1.32

3.6 An application (Hip dataset)

In this section, we implement the Jackson model using both MCMC and INLA for the Hip dataset. We already introduced this dataset in Section 2.4. The main difference between the Hip dataset and the Smoking dataset is that the former has two four-arm trials. However, because of the flexibility of the Jackson model and both inference methods, MCMC and INLA, which we already explained in Sec 3.4 (and referred their corresponding R-code in Appendix), a network meta-analysis dataset which includes a multi-arm trial with any number of arms can be fitted.

The posterior median and 95 % credible intervals for basic contrasts of the fitted Jackson model for the Hip dataset using MCMC and INLA is shown in Figure 3.5. The marginal posterior densities of heterogeneity and inconsistency variances using MCMC and INLA are shown in Figure 3.6. All estimated parameters show very good agreement. Moreover, estimates of inconsistency parameters are shown in Figure 3.7. Again MCMC and INLA results show very good agreement. From marginal posterior density of heterogeneity variance, we can conclude that there is no evidence which advocates a severe

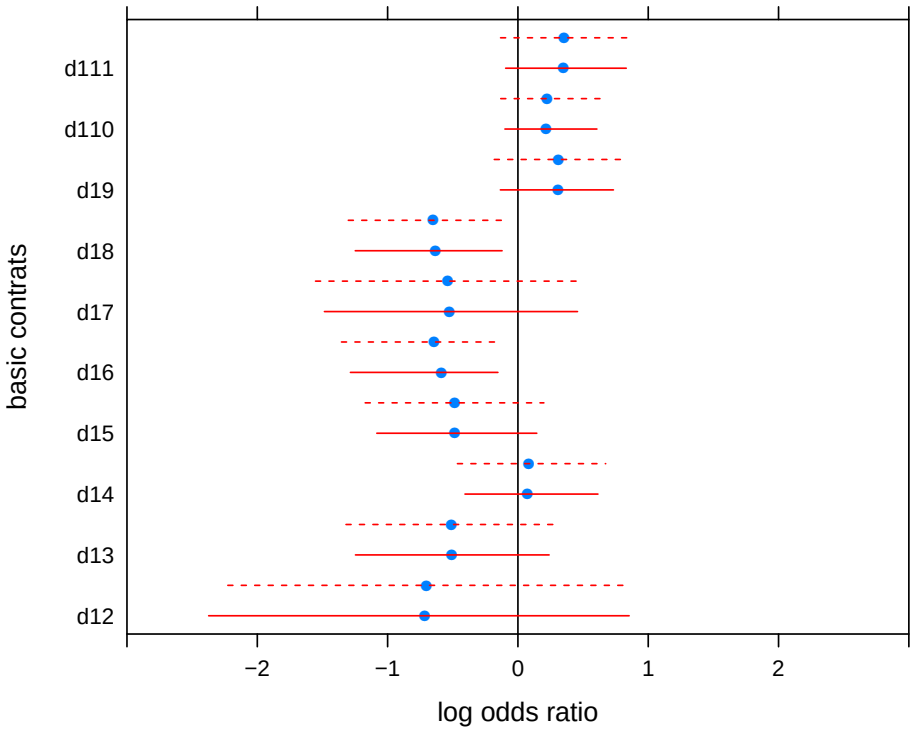

Figure 3.5: Graphical representation of the estimates of the basic contrasts obtained from the implemented Jackson model for the Hip dataset by using MCMC and INLA.

heterogeneity in the network. Also, marginal posterior density of inconsistency variance and estimates of inconsistency parameters both show that there is no evidence suggesting a severe inconsistency in the network. Note that these conclusions are same with the conclusions which we obtained from fitting The Lu-Ades model (see Section 2.4) and also the conclusions of the reference paper (Murad *et al.*, 2012).

R version and packages used to generate this chapter:

R version: R version 3.2.3 (2015-12-10)

Base packages: splines, stats, graphics, grDevices, utils, datasets, methods, base

Other packages: gridExtra, ggplot2, pcnetmeta, nmainla, R2jags, rjags, gemtc, coda, INLA, Matrix, sp, lattice, igraph, xtable

Versions of other packages (respectively): 2.0.0, 2.0.0, 2.3, 1.1, 0.5.7, 4.5, 0.7.1, 0.18.1, 0.0.1443538834, 1.2.3, 1.2.2, 0.20.33, 1.0.1, 1.8.2

This document was generated on February 23, 2016 at 14:14.

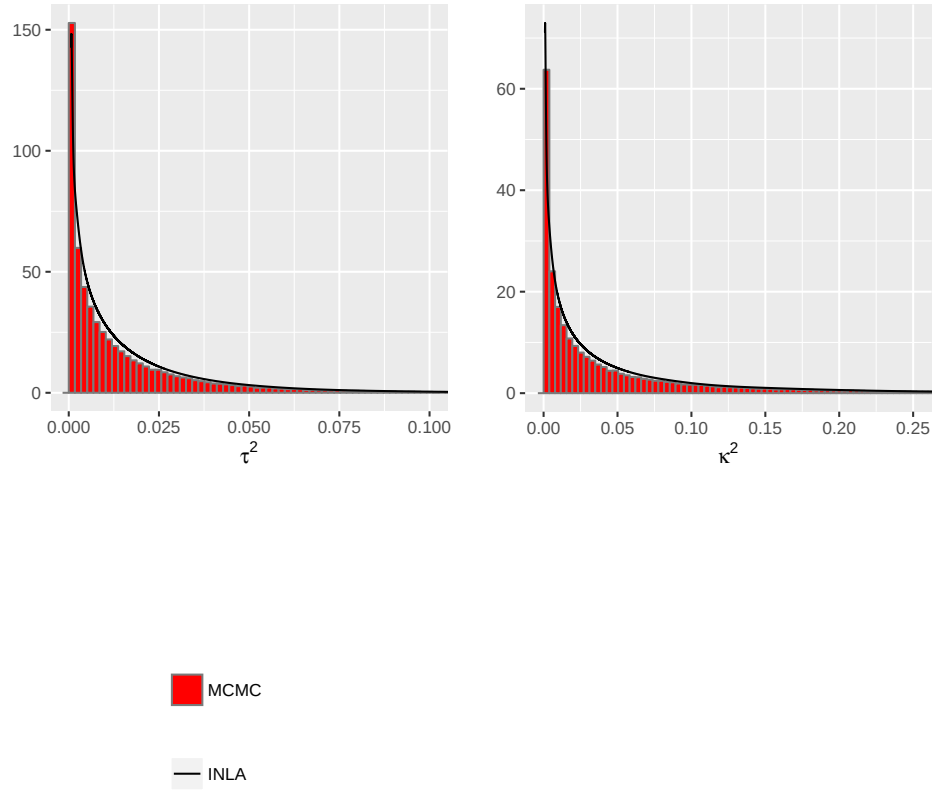

Figure 3.6: Marginal posterior density estimates of the heterogeneity and inconsistency variances by MCMC (histogram) and by INLA (straight line) obtained from the fitted Jackson model for the Hip dataset.

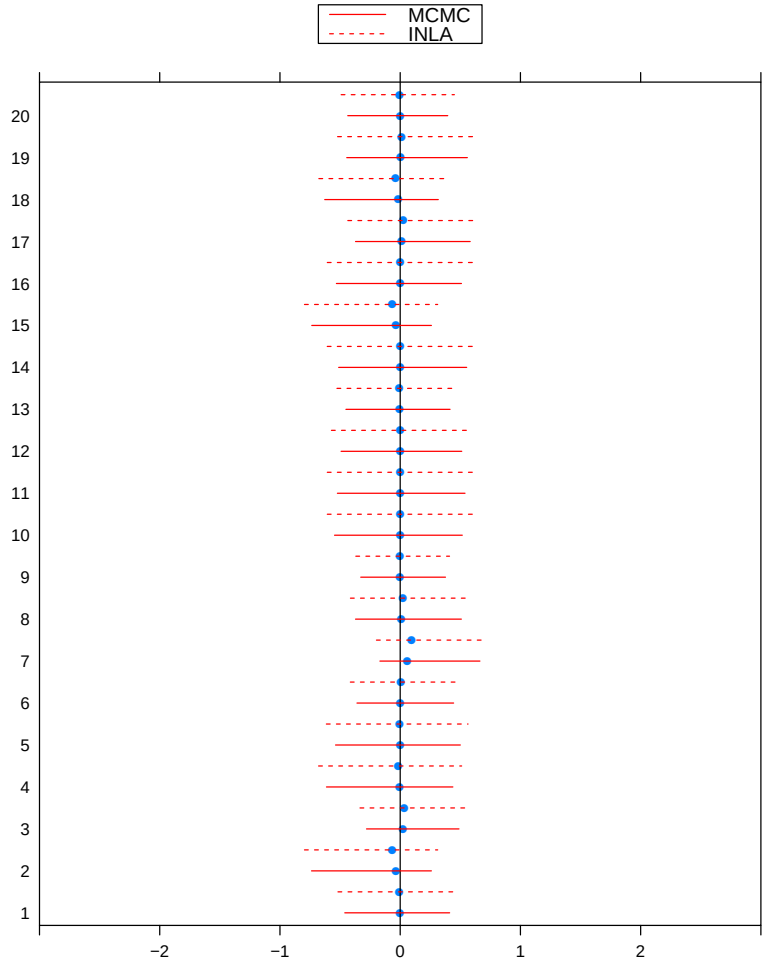

Figure 3.7: Estimates of the inconsistency random effects parameters obtained from the implemented Jackson model for the Hip dataset using MCMC and INLA.

Chapter 4

Conclusions and future research

The thesis project was designed to investigate various inference methods for different statistical models of meta-analysis. With meta-analysis, we mean both pairwise meta-analysis and network meta-analysis. Among different inference methods, an approximate Bayesian inference technique, INLA, was our primary focus.

The application of INLA to pairwise meta-analysis models showed results fairly close to those obtained with its alternatives. Of course, the small differences of the results may stem from INLA's Bayesian nature, since all the alternatives which we discussed in Chapter 1 were either a frequentist or an empirical Bayes method. Moreover, the provided **meta.inla** R function makes the implementing pairwise meta-analysis models (including fixed and random effects models using either a summary level or a trial-arm level with the possibility of meta-regression) with INLA very easy. This function is the first main contribution of the thesis to the current research. With the help of this function, the use of INLA in the pairwise meta-analysis context is as simple as frequentist approaches which is usually one single line code without need to examine any convergence diagnostics. On the other hand, since INLA is an Bayesian technique, the researchers who want to make use of Bayesian methods in this context can enjoy the ease of use of **meta.inla**. Note that the meta-regression option of **meta.inla** can handle up to two covariates. Although, it is possible to include any number of covariates with INLA methodology, this limitation of the function only stem from a programming perspective. Therefore, it is possible to extend the **meta.inla** function to include any number of covariates. Also, one can easily extend the function and make it even more flexible, for example by adding more prior distributions. The fact that the `r-inla` package provides wide options (specifications of different priors, summary methods, model comparison criterion etc.), there are many features which can be included to the **meta.inla** function.

The application of INLA to the consistency model in a NMA context showed results very close to those obtained with MCMC. Note that this conclusion is already made in Sauter and Held (2015). However, in the thesis we apply the INLA method to a dataset including four-arm trials, so our dataset has a more complicated structure than the examples introduced in Sauter and Held (2015). Moreover, we showed that a very

recent statistical model, the Jackson model, with INLA is also possible even in networks with four arm trials. The INLA implementation of the Jackson model using trial-arm level is the second main contribution of the thesis to the current research. Note that INLA (or MCMC) implementation of the Jackson model does not have any limitations regarding the number of arms in the trials of the dataset, so it can handle the dataset including more than four-arm trials. Then, it is also shown that the Lu-Ades models depend on the treatment ordering of the NMA dataset, whereas the Jackson model do not for an application. This task was accomplished by using INLA as an inference method. By relying on these results and the theoretical background which was conceptualized in Jackson *et al.* (2015), we have argued that the design-by-treatment interaction model should be preferred with the presence of the multi-arm trials in the network. From two possible random effects formulations, namely fixed and random effects, of the design-by-treatment interaction model we have argued that the random effects model formulation, the Jackson model, can be preferred because it has easier interpretation compared to the fixed effect formulation. The Jackson model can be used for summary level or a trial-arm level data. A trial arm-level approach has an advantage of not having zero entry problem like summary level. However when there is no arm-level data available, the model using the summary level is the only choice. As far as we understand, INLA implementation of the Jackson model using a summary level is more challenging, if it is possible, a different approach called multivariate meta-regression should be integrated to the INLA methodology.

Our main concern about the fitted NMA models using both MCMC and INLA is the comparison of those two methods. Although, we did not explicitly discuss the computation time with INLA compared to MCMC for our applications, INLA is known for its speed compared to MCMC, mainly because it is not a simulation based technique. On the other hand, we can argue that the key advantage of INLA for NMA models is that there is no need to examine the convergence diagnostics of samples. When implementing MCMC methods, one should be careful about convergence diagnostics as we did in the thesis. Also, this requires some background knowledge in MCMC techniques. From the INLA perspective, implementing NMA models under `r-inla` as we demonstrated is also not straightforward. For example, the use of correlated random effects for multi-arm trials is one of the specialities which we discussed. However, INLA implementation of NMA models can be automated as we achieved by **meta.inla** in the pairwise meta-analysis context. Taking everything into consideration, INLA has a great potential for performing Bayesian inference for both pairwise meta-analysis and NMA models.

There are many possible future research topics regarding the thesis. Writing an R function to make graphical representation of data and **meta.inla** results is one possibility in pairwise meta-analysis context. For NMA, also an **R** function can be created to implement the consistency and the Jackson models (also the Lu-Ades model for networks which include only two-arm trials). By this way, implementing those relatively

complicated models with INLA become available for a wide people of researchers. Also, implementing the Jackson model using summary level is another possible work. As discussed in Jackson *et al.* (2014), regression type models for NMA, network meta-regression, is also possible using the Jackson model. On the other hand we have showed that pairwise meta-regression models can be implemented by using INLA. Therefore extending the Jackson model to fit network meta-regression models with INLA can be considered as a future topic. Lastly, the R functions which we have just described can be included to the `nmainla` R package to make them more available to the researchers.

Appendix A

R-code and BUGS/JAGS-code

A.1 Pairwise meta-analysis models

A.1.1 Random effects model with the method of moment approach

```
# Loading the TB dataset (raw data)
TB <- read.csv("../data/TB.txt")[-1]
N <- nrow(TB)  # number of trials
# Calculating log odds ratios and variances from data
logodds <- function(x) log((x[1] * (x[4] - x[3]))/((x[2] - x[1]) *
    x[3]))
vars <- function(x) 1/x[1] + 1/(x[2] - x[1]) + 1/x[3] + 1/(x[4] -
    x[3])
Y <- apply(cbind(TB$TRTTB, TB$TRT, TB$CONTB, TB$CON), 1,
    logodds)
sigma <- apply(cbind(TB$TRTTB, TB$TRT, TB$CONTB, TB$CON),
    1, vars)
prec <- 1/sigma
# Implementing the moment-based approach
library(metafor)
mom.re.tb <- rma.uni(yi = Y, vi = sigma, method = "DL")
# Mean treatment effect
nu_mom  <- mom.re.tb$b
nuL_mom <- mom.re.tb$ci.lb
nuU_mom <- mom.re.tb$ci.ub
# Heterogeneity variance
tau2_mom <- mom.re.tb$tau2
```

A.1.2 Random effects model with the likelihood approach

```
# Data preparation in order to fit by 'lme' function
d <- rep(1, times = N)  # treatment
ind <- 1:N  # ID
```

```
TB_sum <- data.frame(cbind(d, Y, sigma, ind))
# Likelihood approach for random effects model
library(nlme)
reml.re.tb <- lme(Y ~ -1 + d, random = ~1 | ind, data = TB_sum,
    weights = varConstPower(form = ~sigma, fixed = list(power = 1)),
    method = "REML", na.action = na.omit, control = list(opt = "optim"))
# Mean treatment effect
nu_reml <- fixef(reml.re.tb)
nuL_reml <- nu_reml - 1.96 * summary(reml.re.tb)$tTable[2]
nuU_reml <- nu_reml + 1.96 * summary(reml.re.tb)$tTable[2]
# Heterogeneity variance
tau2_reml <- as.numeric(VarCorr(reml.re.tb)[1, 1])
```

A.1.3 Random effects model with the empirical Bayes approach

```
# confidence level
confLevel <- 0.95
## small positive value
eps <- sqrt(.Machine$double.eps)
# Implementing Empirical Bayes method for the random effects
# model This R-code is mainly adapted from R-code which is
# shown in lecture STA 422 Bayesian Inference(spring 2015) by
# Prof. Leonhard Held

# Defining the function which computes the mean treatment
# effect MLE for fixed heterogeneity variance
nuMle <- function(tau2) {
    precisions <- 1/(sigma + tau2)
    weighted.mean(Y, precisions)
}
## Compute the profile log-likelihood of heterogeneity
## variance
profilLogLikTau2 <- function(tau2) {
    margVars <- sigma + tau2
    arg <- log(margVars) + (Y - nuMle(tau2))^2/margVars
    -1/2 * sum(arg)
}
# Computing by numerical maximization of the profile
# log-likelihood function The empirical Bayes estimate of the
# heterogeneity variance
tau2Ml <- optimize(profilLogLikTau2, c(eps, 1/eps), maximum = TRUE)$maximum
# The empirical Bayes estimate of the mean treatment effect
nuMl <- nuMle(tau2Ml)
# Computing the empirical Bayes estimates for individual
# treatment effects and their corresponding credible
# intervals
weights <- sigma/(sigma + tau2Ml)
postExpectation <- weights * nuMl + (1 - weights) * Y
```

```
postSd <- sqrt((1 - weights) * sigma)
postLower <- postExpectation - postSd * 1.96
postUpper <- postExpectation + postSd * 1.96
# Define functions in order to estimate profile
# log-likelihood confidence interval for the mean treatment
# effect
tau2Mle <- function(nu) {
    scoreTau2 <- function(tau2) {
        margVars <- sigma + tau2
        arg <- (Y - nu)^2/margVars - 1
        arg <- arg/margVars
        1/2 * sum(arg)
    }
    res <- uniroot(scoreTau2, c(eps, 1/eps))
    return(res$root)
}
profilLogLikNu <- function(nu) {
    margVars <- sigma + tau2Mle(nu)
    arg <- log(margVars) + (Y - nu)^2/margVars
    -1/2 * sum(arg)
}
normProfilLogLikNu <- function(nu) profilLogLikNu(nu) - profilLogLikNu(nuMl)
# The function 'likelihood.ci' is directly taken from the
# book Applied Statistical Inference written by Held, L. and
# Sabanes Bove, D.(2014), page numbers are 108 and 109.
# Define a general function which computes likelihood
# confidence intervals
likelihood.ci <- function(gamma, loglik, theta.hat, lower, upper,
    comp.lower = TRUE, comp.upper = TRUE, ...) ## additional arguments for the log-likelihood
{
    ## target function, such that f(theta)=0 gives CI limits
    f <- function(theta, ...) {
        loglik(theta, ...) - loglik(theta.hat, ...) + 1/2 * qchisq(gamma,
            df = 1)
    }
    ## compute lower and upper bounds of CI
    ret <- c()
    if (comp.lower) {
        hl.lower <- uniroot(f, interval = c(lower, theta.hat),
            ...)$root
        ret <- c(ret, hl.lower)
    }
    if (comp.upper) {
        hl.upper <- uniroot(f, interval = c(theta.hat, upper),
            ...)$root
        ret <- c(ret, hl.upper)
    }
    return(ret)
```

```
}
# The 95 % profile likelihood confidence interval for the
# mean treatment effect
likCiNu <- likelihood.ci(gamma = confLevel, loglik = normProfilLogLikNu,
    theta.hat = nuMl, lower = -10, upper = +10)
# Plotting the Figure 1.2 from Chapter 1
library(lattice)  # needed for plotting figure by using 'dotplot' function
# Settings for 'dotplot' function
trials <- c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10",
    "11", "12", "13")
trials_overall <- c(trials, "Random effects model")
trials_overall <- ordered(trials_overall, levels = trials_overall)
ests <- c(postExpectation, nuMl)
ci_Lower <- c(postLower, likCiNu[1])
ci_Upper <- c(postUpper, likCiNu[2])
dotplot(trials_overall ~ ests, xlab = "log odds ratio", ylab = "trial",
    xlim = c(-3, 3), panel = function(x, y) {
        panel.xyplot(x, y, pch = 16, cex = 0.8)
        panel.abline(v = 0, col = 2, lty = 2)
        panel.abline(h = 13.5, col = 1, lty = 1)
        panel.segments(ci_Lower, as.numeric(y), ci_Upper, as.numeric(y),
            lty = 1, col = "red")
    })
```

A.1.4 creatINLAdat.dir and meta.inla R functions

```
# The data preparation function needed for 'meta.inla'
# function
creatINLAdat.dir <- function(ntrt, nctrl, ptrt, pctrl, cov1 = NULL,
    cov2 = NULL) {
    # Adding 0.5 to entries which have 0 value!
    zerocell <- function(y) {
        if (y["ptrt"] == 0) {
            y["ptrt"] <- 0.5
            y["ntrt"] <- y["ntrt"] + 1
        } else if (y["pctrl"] == 0) {
            y["pctrl"] <- 0.5
            y["ntrt"] <- y["ntrt"] + 1
        }
        return(y)
    }
    data <- NULL
    data$ptrt <- ptrt
    data$ntrt <- ntrt
    data$pctrl <- pctrl
    data$nctrl <- nctrl
    data <- data.frame(data)
    N <- nrow(data)  # number of trials
```

```
    data.nozero <- t(apply(data, 1, zerocell))
    d <- rep(1, times = N)  # treatment arm
    Y <- apply(data.nozero, 1, function(x) log((x[1] * (x[4] -
        x[3]))/((x[2] - x[1]) * x[3])))  # observed log odds ratios
    prec <- 1/apply(data.nozero, 1, function(x) 1/x[1] + 1/(x[2] -
        x[1]) + 1/x[3] + 1/(x[4] - x[3]))  # precisions
    re <- 1:nrow(data.nozero)  # ID for random effects
    data.sum <- data.frame(cbind(d, Y, prec, re, cov1, cov2))
    Y <- as.vector(rbind(data$pctrl, data$ptrt))  # number of diseased patients
    sampleSize <- as.vector(rbind(data$nctrl, data$ntrt))  # number of all patients
    d <- rep(0:1, times = N)
    re <- as.vector(rbind(rep(NA, times = N), 1:N))  # heterogeneity random effects
    if (!is.null(cov1)) {
        cov1 <- as.vector(rbind(NA, cov1))
    }
    if (!is.null(cov2)) {
        cov2 <- as.vector(rbind(NA, cov2))
    }
    data.arm <- data.frame(cbind(Y, sampleSize, d, re, cov1,
        cov2))
    data.arm$mu <- as.factor(as.numeric(gl(n = N, k = 2)))
    datINLA <- list(data.sum = data.sum, data.arm = data.arm)
    return(datINLA)
}
# THE 'meta.inla' function
# Pairwise meta-analysis with INLA
meta.inla <- function(datINLA, meanf = 0, varf = 1000, ul = 10,
    digits = 3, mod = "FE", type = "summary", mreg = FALSE, ...)
{
    if (mod %in% c("FE", "RE") == FALSE) {
        stop("Function argument \"mod\" must be equal to \"FE\" or \"RE\"!")
    }
    if (type %in% c("summary", "trial-arm") == FALSE) {
        stop("Function argument \"type\" must be equal to \"summary\" or \"trial-arm\"!")
    }
    if (mreg == TRUE && is.null(datINLA$data.sum$cov1)) {
        stop("Function argument \"cov1\" must not be equal to \"NULL\" !")
    }
    # Prior dist for hyperparameter --Function for Uniform
    # distribution:
    hyperunif.function <- function(x) {
        if (exp(x)^-0.5 < ul & exp(x)^-0.5 > 0) {
            logdens <- log(1/ul)
        } else {
            logdens <- log(0.1e-320)
        }
        logdenst <- logdens + log(0.5 * exp(-x/2))
        return(logdenst)
```

```
}
lprec <- seq(from = -40, to = 40, len = 20000)
# Create table:
prior.table <- paste(c("table:", cbind(lprec, sapply(lprec,
    FUN = hyperunif.function))), sep = "", collapse = " ")
inla.form <- "Y ~ -1 + d"
if (type == "trial-arm") {
    inla.form <- paste(inla.form, " + mu ", sep = "")
}
if (mreg == TRUE) {
    inla.form <- paste(inla.form, " + cov1 ", sep = "")
}
if (mreg == TRUE && !is.null(datINLA$data.sum$cov2)) {
    inla.form <- paste(inla.form, " + cov2 ", sep = "")
}
if (mod == "RE") {
    inla.form <- paste(inla.form, " + f(re, model=\"iid\", hyper = list(theta =
                    list(prior = prior.table))", ")", sep = "")
}
if (type == "summary") {
    fit.inla <- inla(as.formula(inla.form), data = datINLA$data.sum,
        family = "normal", control.fixed = list(expand.factor.strategy = "inla",
            mean = meanf, prec = 1/varf), control.family = list(hyper = list(prec =
            list(fixed = TRUE, initial = 0))), scale = prec, ...)
}
if (type == "trial-arm") {
    fit.inla <- inla(as.formula(inla.form), data = datINLA$data.arm,
        family = "binomial", control.fixed = list(expand.factor.strategy = "inla",
            mean = 0, prec = 1/varf), Ntrials = sampleSize,
        ...)
}
nu <- as.numeric(fit.inla$summary.fixed[1, c(3, 4, 5)])
tau2 <- rev(1/summary(fit.inla)$hyperpar[1, c(3, 4, 5)])
if (mreg == TRUE) {
    if (type == "trial-arm") {
        N <- nrow(datINLA$data.arm)/2
    } else N <- 0
    cov <- as.numeric(fit.inla$summary.fixed[N + 2, c(3,
        4, 5)])
    if (mreg == TRUE && !is.null(datINLA$data.sum$cov2)) {
        inla.form <- paste(inla.form, " + cov2 ", sep = "")
        results <- list(nu = nu, cov = cov, tau2 = tau2,
            call = match.call(), mreg = mreg, mod = mod)
    }
    cov2 <- as.numeric(fit.inla$summary.fixed[N + 3, c(3,
        4, 5)])
    results <- list(nu = nu, cov = cov, cov2 = cov2, tau2 = tau2,
        call = match.call(), mreg = mreg, mod = mod)
```

```
    } else results <- list(nu = nu, tau2 = tau2, call = match.call(),
        mreg = mreg, mod = mod)
    class(results) <- "meta.inla"
    return(results)
}
```

A.1.5 Conducting meta-analysis with meta.inla function

```
# Data pareparation --needed for meta.inla function
TB.datINLA <- creatINLAdat.dir(ntrt = TB$TRT, nctrl = TB$CON,
    ptrt = TB$TRTTB, pctrl = TB$CONTB)
# Fitting fixed effect model using summary level approach
library(INLA)
inla.fe.tb <- meta.inla(TB.datINLA, varf = 1000, mod = "FE",
    type = "summary")
# Fitting random effects model using summary level approach
inla.re.tb <- meta.inla(TB.datINLA, varf = 1000, ul = 10, mod = "RE",
    type = "summary")
# Fitting fixed effect model using trial-arm level approach
inla.fe.arm.tb <- meta.inla(TB.datINLA, varf = 1000, mod = "FE",
    type = "trial-arm")
# Fitting random effects model using trial-arm level approach
inla.re.arm.tb <- meta.inla(TB.datINLA, varf = 1000, ul = 10,
    mod = "RE", type = "trial-arm")
```

A.1.6 Random effects meta-regression with weighted-least squares and mixed model approaches

```
# Centering the covariates about the study number '5' to aid
# interpretation
TB$Year_centered <- TB$Year - TB$Year[5]
TB$Latitude_centered <- TB$Latitude - TB$Latitude[5]
# Using wlsq approach
wlsq.mreg.tb <- rma.uni(yi = Y, vi = sigma, mods = Y ~ TB$Latitude_centered +
    TB$Year_centered, method = "DL")
# Using reml approach
TB_sum$Year_centered <- TB$Year_centered
TB_sum$Latitude_centered <- TB$Latitude_centered
nlme.mreg.tb <- lme(Y ~ -1 + d + Latitude_centered + Year_centered,
    random = ~1 | ind, data = TB_sum, weights = varConstPower(form = ~sigma,
        fixed = list(power = 1)), method = "REML", na.action = na.omit,
    control = list(opt = "optim"))
# Heterogeneity variance
tau2_like <- as.numeric(VarCorr(nlme.mreg.tb)[1, 1])
```

A.1.7 Conducting meta-regression with meta.inla function

```
# Creating data frame by including covariates
datINLA <- creatINLAdat.dir(ntrt = TB$TRT, nctrl = TB$CON,
    ptrt = TB$TRTTB, pctrl = TB$CONTB, cov1 = TB$Latitude_centered,
    cov2 = TB$Year_centered)
# Fixed effect meta regression using summary level
inla.mreg.fe.tb <- meta.inla(datINLA, mod = "FE", type = "summary",
    mreg = TRUE)
# Random effects meta regression using summary level
inla.mreg.re.tb <- meta.inla(datINLA, mod = "RE", type = "summary",
    mreg = TRUE)
# Fixed effect meta regression using trial-arm level
inla.mreg.fe.arm.tb <- meta.inla(datINLA, mod = "FE", type = "trial-arm",
    mreg = TRUE)
# Random effects meta regression using trial-arm level
inla.mreg.fe.arm.tb <- meta.inla(datINLA, mod = "RE", type = "trial-arm",
    mreg = TRUE)
```

A.2 The consistency model

A.2.1 Data preparation and data visualization for an NMA

```
# Load 'nmainla' R-package
library(nmainla)
# Load the Hip dataset
hip_fracdat_raw <- read.csv("../data/HipFracture.txt", header = T, sep = "\t")
# The necessary operation --switch names treatment 10 and 1!
# This is needed becaues of technical reasons caused by
# 'creatINLAdat' function: Global baseline treatment should
# be treatment 1, so we should switch coding of treatment 10
# and treatment 1!
for (k in 1:4) {
    tk <- paste("t", k, sep = "")
    hip_fracdat_raw[[tk]][which(hip_fracdat_raw[[tk]] == 1)] <- 0
    hip_fracdat_raw[[tk]][which(hip_fracdat_raw[[tk]] == 10)] <- 1
    hip_fracdat_raw[[tk]][which(hip_fracdat_raw[[tk]] == 0)] <- 10
}
# Change data-format to one-arm-per-row data
hip_fracdat <- mtc.data.studyrow(data = hip_fracdat_raw, armVars = c(treatment = "t",
    responders = "r", sampleSize = "n"), nArmsVar = "NA.", studyNames = 1:nrow(hip_fracdat_raw),
    patterns = c("%s", "%s%d"))
# Plot network:
library(pcnetmeta)
nma.networkplot(s.id = study, t.id = treatment, data = hip_fracdat)
```

A.2.2 Fitting the consistency model using **JAGS** with `R2jags`

```
# Settings needed for MCMC
niter <- 6e+05  # number of iterations
burnin <- 3e+05  # number of burnin
thin <- 3  # number of thinning
nchain <- 3  # number of MCMC chains
# Dataset
jags.hip <- list(nt = 11, ns = 40, r = cbind(hip_fracdat_raw$r1,
    hip_fracdat_raw$r2, hip_fracdat_raw$r3, hip_fracdat_raw$r4),
    n = cbind(hip_fracdat_raw$n1, hip_fracdat_raw$n2, hip_fracdat_raw$n3,
        hip_fracdat_raw$n4), t = cbind(hip_fracdat_raw$t1, hip_fracdat_raw$t2,
        hip_fracdat_raw$t3, hip_fracdat_raw$t4), na = hip_fracdat_raw$NA.,
    prec = 1/varf, ul = ul)
# Parameters to save
params.hip <- c("delta", "taubeta")
# Run JAGS model:
jags.hip.RE.JAGS <- jags(model = "JAGS.Cons.Model.txt", parameters.to.save = params.hip,
    data = jags.hip, n.chains = nchain, burnin = burnin, n.iter = niter,
    n.thin = thin)
# Convergence diagnostics and results
plot(jags.hip.RE.JAGS)
summary(jags.hip.RE.JAGS)
```

A.2.3 Fitting the consistency model using `r-inla`

```
# Prior specifications
varf <- 1000
ul <- 10
# Function for Uniform distribution:
hyperunif.function <- function(x) {
    if (exp(x)^-0.5 < ul & exp(x)^-0.5 > 0) {
        logdens <- log(1/ul)
    } else {
        logdens <- log(9.98012604599318e-322)
    }
    logdenst <- logdens + log(0.5 * exp(-x/2))
    return(logdenst)
}
# Set up grid to evaluate the uniform prior:
lprec <- seq(from = -40, to = 40, len = 20000)  ## CHANGE this LINE if INLA crashes!
## (extend grid by changing (from= , to=)) Create table with
## prior values and lprec:
prior.table <- paste(c("table:", cbind(lprec, sapply(lprec, FUN = hyperunif.function))),
    sep = "", collapse = " ")
# Some more data preparation steps for INLA approach Creating
# Baseline variable:
hip_fracdat_raw$bas <- apply(hip_fracdat_raw[, c("t1", "t2",
    "t3", "t4")], 1, function(x) min(x, na.rm = TRUE))
hip_fracdat$baseline <- rep(hip_fracdat_raw$bas, times = hip_fracdat_raw$NA.)
```

```
model{

# LOOP THROUGH STUDIES
for(i in 1:ns){

# BINOMIAL LIKELIHOOD WITH LOGIT LINK
# VAGUE PRIORS FOR TRIAL BASELINES
    base[i] ~ dnorm(0, prec)
# LOOP THROUGH ARMS
    for (k in 1:na[i])  {
         r[i,k] ~ dbin(p[i,k],n[i,k])
         logit(p[i,k]) <- base[i] + eta[i,k]
       }
# RANDOM EFFECTS DISTRIBUTION
w[i,1] <- 0
    eta[i,1] <- 0
   for (k in 2:na[i]) {  # LOOP THROUGH ARMS
        eta[i,k] ~ dnorm(m.cond[i,k], precbeta.cond[i,k])
# MEANS WITH MULTI-ARM TRIAL CORRECTION
        m.cond[i,k] <-  delta[t[i,k]] - delta[t[i,1]] + sw[i,k]
# BETWEEN-STUDY PRECISION WITH MULTI-ARM TRIAL CORRECTION
        precbeta.cond[i,k] <- precbeta * 2 * (k-1)/k
        w[i,k] <- (eta[i,k] - delta[t[i,k]] + delta[t[i,1]])
        sw[i,k] <- sum(w[i,1:(k-1)]) / (k-1)
      }
  }
# TREATMENT EFFECT IS ZERO FOR REFERENCE TREATMENT
delta[1] <- 0
# VAGUE PRIORS
for (k in 2:nt) { delta[k] ~ dnorm(0,prec) }
taubeta  ~  dunif(0,ul)
precbeta <- pow(taubeta,-2)

}
```

Listing A.1: BUGS/JAGS code for the consistency model (JAGS.Cons.Model).

```
# Study should be factor variable!
hip_fracdat$mu <- as.factor(hip_fracdat$study)
# Creating data frame suitable for 'inla'
hipfracdatINLA <- creatINLAdat(dat = hip_fracdat, treatmentvar = "treatment",
    baselinevar = "baseline", studyvar = "study")
# Some Settings Accounting for multi-arm trials Transform
# group-correlation 0.5 to internal.scale of INLA:
cor <- 0.5  # correlation between treatment comparisons of the same multi-arm trial.
ngroup <- 3  # number of groups is equal to the maximum number
# of pairwise treatment comparisons in a (multi-arm) trial
# transformation to internal INLA-scale.
```

```
cor.inla.init <- log((1 + cor * (ngroup - 1))/(1 - cor))
# The Consistency model
inla_hip_arm_form_RE <- responders ~ -1 + mu + d12 + d13 + d14 +
    d15 + d16 + d17 + d18 + d19 + d110 + d111 + f(re, model = "iid",
    hyper = list(theta1 = list(prior = prior.table)), group = g,
    control.group = list(model = "exchangeable", hyper = list(rho = list(fixed = TRUE,
        initial = cor.inla.init))))
# Call inla:
nmainlaRE.hip <- inla(as.formula(inla_hip_arm_form_RE), Ntrials = sampleSize,
    family = "binomial", data = hipfracdatINLA, control.fixed =
            list(expand.factor.strategy = "inla",
        mean = 0, prec = 1/varf), control.compute = list(dic = TRUE,
        cpo = TRUE), control.inla = list(strategy = "simplified.laplace",
        lincomb.derived.only = FALSE))
nmainlaRE.hip <- inla.hyperpar(nmainlaRE.hip)
# Basic contrasts
inla.bas.hip <- nmainlaRE.hip$summary.fixed[41:51, ][, c(3, 4,
    5)]
# Heterogeneity standard deviation
tau2.inla <- rev(sqrt(1/summary(nmainlaRE.hip)$hyperpar[1, -c(1,
    2, 6)]))
```

A.3 The Lu-Ades model

A.3.1 Fitting inconsistency model using cycle-specific approach with r-inla

```
# Implementation of the Lu-Ades model Special attention to
# cycle-specific inconsistency random effects --ICDF must be
# determined by 'hand'!
hipfracdatINLA$w <- rep(NA, nrow(hipfracdatINLA))
hipfracdatINLA[hipfracdatINLA$study == 20 & hipfracdatINLA$treatment ==
    8, ]$w <- 1
hipfracdatINLA[hipfracdatINLA$study == 1 & hipfracdatINLA$treatment ==
    10, ]$w <- 2
hipfracdatINLA[hipfracdatINLA$study == 2 & hipfracdatINLA$treatment ==
    10, ]$w <- 3
hipfracdatINLA[hipfracdatINLA$study == 3 & hipfracdatINLA$treatment ==
    10, ]$w <- 3
hipfracdatINLA[hipfracdatINLA$study == 29 & hipfracdatINLA$treatment ==
    9, ]$w <- 4
hipfracdatINLA[hipfracdatINLA$study == 37 & hipfracdatINLA$treatment ==
    11, ]$w <- 5
hipfracdatINLA[hipfracdatINLA$study == 4 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
hipfracdatINLA[hipfracdatINLA$study == 5 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
```

```
hipfracdatINLA[hipfracdatINLA$study == 6 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
hipfracdatINLA[hipfracdatINLA$study == 7 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
hipfracdatINLA[hipfracdatINLA$study == 8 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
hipfracdatINLA[hipfracdatINLA$study == 9 & hipfracdatINLA$treatment ==
    10, ]$w <- 6
# Inconsitency random effects
lc1 <- inla.make.lincomb(d14 = -1, d16 = 0, d18 = 1, d19 = 0,
    d110 = 0, d111 = 0, w = c(1, NA, NA, NA, NA, NA))
lc2 <- inla.make.lincomb(d14 = -1, d16 = 0, d18 = 0, d19 = 0,
    d110 = 1, d111 = 0, w = c(NA, 1, NA, NA, NA, NA))
lc3 <- inla.make.lincomb(d14 = 0, d16 = -1, d18 = 0, d19 = 0,
    d110 = 1, d111 = 0, w = c(NA, NA, 1, NA, NA, NA))
lc4 <- inla.make.lincomb(d14 = 0, d16 = 0, d18 = -1, d19 = 1,
    d110 = 0, d111 = 0, w = c(NA, NA, NA, 1, NA, NA))
lc5 <- inla.make.lincomb(d14 = 0, d16 = 0, d18 = -1, d19 = 0,
    d110 = 0, d111 = 1, w = c(NA, NA, NA, NA, 1, NA))
lc6 <- inla.make.lincomb(d14 = 0, d16 = 0, d18 = 0, d19 = -1,
    d110 = 1, d111 = 0, w = c(NA, NA, NA, NA, NA, 1))
LC <- list()
LC[["d48"]] <- lc1[[1]]
LC[["d410"]] <- lc2[[1]]
LC[["d610"]] <- lc3[[1]]
LC[["d89"]] <- lc4[[1]]
LC[["d811"]] <- lc5[[1]]
LC[["d910"]] <- lc6[[1]]
# Formula:
inla_hip_arm_form_REinc <- responders ~ -1 + mu + d12 + d13 +
    d14 + d15 + d16 + d17 + d18 + d19 + d110 + d111 + f(re, model = "iid",
    hyper = list(theta1 = list(prior = prior.table)), group = g,
    control.group = list(model = "exchangeable", hyper = list(rho = list(fixed = TRUE,
        initial = cor.inla.init)))) + f(w, model = "iid",
        hyper = list(theta1 = list(prior = prior.table)))
# Call inla:
nmainla.hip <- inla(as.formula(inla_hip_arm_form_REinc), Ntrials = sampleSize,
    family = "binomial", data = hipfracdatINLA, lincomb = LC,
    control.fixed = list(expand.factor.strategy = "inla", mean = 0,
        prec = 1/varf), control.compute = list(dic = TRUE, cpo = TRUE),
    control.inla = list(strategy = "simplified.laplace", lincomb.derived.only = FALSE))
nmainla.hip <- inla.hyperpar(nmainla.hip)
# Results
results.inla <- summary(nmainla.hip)
# Basic contrasts
INLA.luades <- nmainla.hip$summary.fixed[41:50, ][, c(3, 4, 5)]
# Functional contrasts
INLA.luades.fun <- nmainla.hip$summary.lincomb[, c(4, 5, 6)]
```

```
# Get median estimates of marginal posterior of the
# hyperparameter variances: Heterogeneity
het.luades <- rev(sqrt(1/summary(nmainla.hip)$hyperpar[1, -c(1,
    2, 6)]))
# Inconsistency
inc.luades <- rev(sqrt(1/summary(nmainla.hip)$hyperpar[2, -c(1,
    2, 6)]))
# Inconsistency random effects
INLA.luades.incs <- nmainla.hip$summary.random$w
```

A.4 The Jackson model

A.4.1 Fitting Jackson model with JAGS

```
# Load the Smoking dataset
data("smokdatDI", package = "nmainla")
# Adding 'design' variable to the dataset --by hand
smokdatDI$des <- c("1", "2", rep("3", times = 7), "4", "4", "3",
    "3", "3", "5", "4", rep("3", times = 4), "6", "7", "8", "8")
# MCMC settings
niter <- (1e+05 + 1e+05)  # number of iterations
burnin <- 1e+05  # number of burnin
thin <- 5  # number of thinning
nchain <- 3  # number of MCMC chains
# Dataset
jags.smoking.inc <- list(nt = 4, ns = 24, ndes = 8, nades = c(3,
    3, 2, 2, 2, 2, 2, 2), r = cbind(smokdatDI$r1, smokdatDI$r2,
    smokdatDI$r3), n = cbind(smokdatDI$n1, smokdatDI$n2, smokdatDI$n3),
    t = cbind(smokdatDI$t1, smokdatDI$t2, smokdatDI$t3), na = smokdatDI$na,
    des = smokdatDI$des, prec = 1/varf, ul = ul)
params.smoking.inc <- c("delta[2]", "delta[3]", "delta[4]", "taubeta",
    "tauomega", "om[1,2]", "om[1,3]", "om[2,2]", "om[2,3]", "om[3,2]",
    "om[4,2]", "om[5,2]", "om[6,2]", "om[7,2]", "om[8,2]")
# Run JAGS model:
jags.smokeREinc.JAGS <- jags(data = jags.smoking.inc, n.iter = niter,
    n.burnin = burnin, n.thin = thin, n.chains = nchain,
    parameters.to.save = params.smoking.inc, model.file = "JAGS.Design.Model.txt")
# Convergence diagnostics and results
plot(jags.smokeREinc.JAGS)
print(jags.smokeREinc.JAGS)
```

A.4.2 Fitting Jackson model with r-inla

```
data("smokdatDI", package = "nmainla")
# Adding 'design' variable to the dataset --by hand
smokdatDI$des <- c("1", "2", rep("3", times = 7), "4", "4", "3",
    "3", "3", "5", "4", rep("3", times = 4), "6", "7", "8", "8")
```

```
# Data preparation for the Smoking dataset
smokdat <- mtc.data.studyrow(data = smokdatDI, armVars = c(treatment = "t",
    responders = "r", sampleSize = "n"), nArmsVar = "na", studyNames = 1:nrow(smokdatDI),
    patterns = c("%s", "%s%d"))
smokdat$baseline <- rep(smokdatDI$t1, times = smokdatDI$na)
smokdat$mu <- as.factor(smokdat$study)
smokdatINLA <- creatINLAdat(dat = smokdat, treatmentvar = "treatment",
    baselinevar = "baseline", studyvar = "study")
# Adding indicator variable for design inconsistency
# parameters (des) --design inconsistency random effects
smokdatINLA$des <- rep(smokdatDI$des, times = smokdatDI$na)
for (i in 1:nrow(smokdatINLA)) {
    if (smokdatINLA$re[i] %in% NA)
        smokdatINLA$des[i] <- NA
}
# Formula:
inla_form.smokeDesinc <- responders ~ -1 + mu + d12 + d13 + d14 +
    f(re, model = "iid", hyper = list(theta1 = list(prior = prior.table)),
        group = g, control.group = list(model = "exchangeable",
            hyper = list(rho = list(fixed = TRUE, initial = cor.inla.init)))) +
    f(des, model = "iid", hyper = list(theta1 = list(prior = prior.table)),
        group = g, control.group = list(model = "exchangeable",
            hyper = list(rho = list(fixed = TRUE, initial = cor.inla.init))))
# Call inla:
inla.smokeDesinc <- inla(as.formula(inla_form.smokeDesinc), Ntrials = sampleSize,
    family = "binomial", data = smokdatINLA,
    control.fixed = list(expand.factor.strategy = "inla",
        mean = 0, prec = 1/varf), control.compute = list(dic = TRUE,
        cpo = TRUE), control.inla = list(strategy = "simplified.laplace",
        lincomb.derived.only = FALSE))
inla.smokeDesinc <- inla.hyperpar(inla.smokeDesinc)
# Results Get marginal posterior of the hyperparameter
# variances (instead of the precisions):
tau.inla.des <- rev(sqrt(1/summary(inla.smokeDesinc)$hyperpar[1,
    -c(1, 2, 6)]))
inc.inla.des <- rev(sqrt(1/summary(inla.smokeDesinc)$hyperpar[2,
    -c(1, 2, 6)]))
```

```
model{

# LOOP THROUGH STUDIES
for(i in 1:ns){

# BINOMIAL LIKELIHOOD WITH LOGIT LINK
# VAGUE PRIORS FOR TRIAL BASELINES
    base[i] ~ dnorm(0, prec)
# LOOP THROUGH ARMS
    for (k in 1:na[i])  {
         r[i,k] ~ dbin(p[i,k],n[i,k])
         logit(p[i,k]) <- base[i] + eta[i,k] + om[des[i], k]
       }
# RANDOM EFFECTS DISTRIBUTION
w[i,1] <- 0
    eta[i,1] <- 0
   for (k in 2:na[i]) {  # LOOP THROUGH ARMS
        eta[i,k] ~ dnorm(m.cond[i,k], precbeta.cond[i,k])
# MEANS WITH MULTI-ARM TRIAL CORRECTION
        m.cond[i,k] <-  delta[t[i,k]] - delta[t[i,1]] + sw[i,k]
# BETWEEN-STUDY PRECISION WITH MULTI-ARM TRIAL CORRECTION
        precbeta.cond[i,k] <- precbeta * 2 * (k-1)/k
        w[i,k] <- (eta[i,k] - delta[t[i,k]] + delta[t[i,1]])
        sw[i,k] <- sum(w[i,1:(k-1)]) / (k-1)
      }
  }
# INCONSISTENCY PARAMETERS
        for (i in 1:ndes)  {          # LOOP THROUGH DESIGNS
                    om[i,1] <- 0
                    for(k in 2:nades[i])  {  # LOOP THROUGH ARM OF DESIGN i
                        om[i,k] ~ dnorm(mom.cond[i,k],precom.cond[i,k])
# MEAN OF INCONSISTENCY DISTRIBUTION WITH MULTI-ARM TRIAL CORRECTION
                        mom.cond[i,k] <- sum(om[i,1:(k-1)])/(k-1)
# PRECISION OF INCONSISTENCY DISTRIBUTION WITH MULTI-ARM TRIAL CORRECTION
                        precom.cond[i,k] <- precomega * 2 * (k-1)/k
                }
 }
# TREATMENT EFFECT IS ZERO FOR REFERENCE TREATMENT
delta[1] <- 0
# VAGUE PRIORS
for (k in 2:nt) { delta[k] ~ dnorm(0,prec) }
taubeta  ~  dunif(0,ul)
precbeta <- pow(taubeta,-2)
varbeta  <- 1 / precbeta
tauomega  ~  dunif(0,ul)
precomega <- pow(tauomega,-2)
varomega  <- 1 / precomega

}
```

Listing A.2: BUGS/JAGS code for the Jackson model (JAGS.Design.Model).

Acknowledgments

I would like to thank one of my supervisor Prof. Leonhard Held for giving me the opportunity to work on this exiting project and his kind and thorough support as well as his good advice to improve and extend the thesis. Then, I would like to thank Rafael Sauter who is my other supervisor for his support and in-depth discussions about many topics in the thesis but also through useful general guidelines about performing the research. He always had an open ear for all kinds of my questions.

Bibliography

Baker, S. G. and Kramer, B. S. (2002). The transitive fallacy for randomized trials: if a bests b and b bests c in separate trials, is a better than c? *BMC Medical Research Methodology*, **2**, 13. 24

Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2005). Comprehensive meta-analysis version 2. *Englewood, NJ: Biostat*, **104**, . 32

Bravata, D. M. and Olkin, I. (2001). Simple pooling versus combining in meta-analysis. *Evaluation & the health professions*, **24**, 218–230. 4

Chalmers, I., Hedges, L. V., and Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the health professions*, **25**, 12–37. 2

Coldlitz, G., Brewer, T., Berkey, C., Wilson, M., Burdick, E., Fineberg, H., and Mosteller, F. (1994). Efficacy of bcg vaccine in the prevention of tuberculosis. *J. Am. Med. Assoc*, **271**, 698–702. 2

Cox, D. R. and Snell, E. J. (1989). *Analysis of binary data*, volume 32. CRC Press. 33

Davison, A. C. (2003). *Statistical models*, volume 11. Cambridge University Press. 9

Deal, C. L. (1997). Osteoporosis: prevention, diagnosis, and management. *The American journal of medicine*, **102**, 35S–39S. 32

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, **7**, 177–188. 7

Dias, S., Welton, N., Caldwell, D., and Ades, A. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Statistics in medicine*, **29**, 932–944. 14, 28, 41, 47

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press. 8

Everitt, B. S. (2002). *Cambridge Dictionary of Statistics in the Medical Sciences*. UK: Cambridge University Press. 2

Garg, A. X., Hackam, D., and Tonelli, M. (2008). Systematic review and meta-analysis: when one study is just not enough. *Clinical Journal of the American Society of Nephrology*, **3**, 253–260. 2

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, **7**, 457–472. 35

Harvard, T. C. and Lau, J. (1993). Meta-analytic stimulus for changes in clinical trials. *Statistical methods in medical research*, **2**, 161–172. 2

Hasselblad, V. (1998). Meta-analysis of multitreatment studies. *Medical Decision Making*, **18**, 37–43. 47

Held, L. and Sabanés Bové, D. (2014a). *Applied statistical inference*. Springer. 30

Held, L. and Sabanés Bové, D. (2014b). Empirical bayes methods. In *Applied Statistical Inference*, 209–214. Springer. 9

Held, L. and Sabanés Bové, D. (2014c). Posterior distribution. In *Applied Statistical Inference*, 170–179. Springer. 8

Higgins, J., Jackson, D., Barrett, J., Lu, G., Ades, A., and White, I. (2012). Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods*, **3**, 98–110. 31, 41, 42, 44, 46, 47, 51

Higgins, J., Thompson, S. G., and Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **172**, 137–159. 6, 7, 12

Higgins, J. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in medicine*, **15**, 2733–2749. 29

Hothorn, T. and Everitt, B. S. (2014). Systematic review and meta-analysis. In *A handbook of statistical analyses using R*, 267–284. CRC press. 1

Jackson, D., Barrett, J. K., Rice, S., White, I. R., and Higgins, J. (2014). A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Statistics in medicine*, **33**, 3639–3654. 35, 42, 45, 46, 47, 50, 59

Jackson, D., Boddington, P., and White, I. R. (2015). The design-by-treatment interaction model: a unifying framework for modelling loop inconsistency in network meta-analysis. *Research synthesis methods*, ***n***/a–n/a, . 52, 58

Jaki, T. (2015). Multi-arm clinical trials with treatment selection: what can be gained and at what price? *Clinical Investigation*, **5**, 393–399. 23, 29

Last, J., Spasoff, R., Harris, S., and Thuriaux, M. (2001). International epidemiological association. a dictionary of epidemiology. 1

Lau, J., Ioannidis, J. P., and Schmid, C. H. (1998). Summing up evidence: one answer is not always enough. *The lancet*, **351**, 123–127. 15

Lu, G. and Ades, A. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in medicine*, **23**, 3105–3124. 29, 33

Lu, G. and Ades, A. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, **101**, . 25, 26, 27, 28, 31, 41

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in medicine*, **21**, 2313–2324. 7, 24, 27, 41

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, **10**, 325–337. 33

Murad, M. H., Drake, M. T., Mullan, R. J., Mauck, K. F., Stuart, L. M., Lane, M. A., Abu Elnour, N. O., Erwin, P. J., Hazem, A., Puhan, M. A.,*et al.* (2012). Comparative effectiveness of drug treatments to prevent fragility fractures: a systematic review and network meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, **97**, 1871–1880. 31, 35, 54

Parmar, M. K., Carpenter, J., and Sydes, M. R. (2014). More multiarm randomised trials of superiority are needed. *The Lancet*, **384**, 283–284. 29

Petitti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine*. Oxford University Press. 2

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-122. 8

Plummer, M.*et al.* (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, 125. Technische Universit at Wien. 34

Puhan, M. A., Schünemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., Guyatt, G. H.,*et al.* (2014). A grade working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*, **349**, g5630. 23, 24

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. iii

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. *The handbook of research synthesis and meta-analysis*, **2**, 295–316. 16

Rayleigh, L. (1885). Address by the Rt. Hon. Lord Rayleigh. Montreal, Canada: John Murray. Accessed 2015-09-11, http://www.jameslindlibrary.org/rayleigh-the-lord-1885/. 1

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, **71**, 319–392. 12

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research synthesis methods*, **3**, 80–97. 21, 23, 24, 25

Sauter, R. and Held, L. (2015). Network meta-analysis with integrated nested laplace approximations. *Biometrical Journal*, **57**, 1038–1050. 12, 13, 24, 25, 28, 29, 34, 36, 41, 47, 57

Schwarzer, G., Carpenter, J. R., and Rücker, G. (2015). *Meta-analysis with R*. Springer. 16, 22, 25

Song, F., Loke, Y. K., Walsh, T., Glenny, A.-M., Eastwood, A. J., Altman, D. G.,*et al.* (2009). Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *Bmj*, **338**, . 41

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons. 12

Su, Y.-S. and Yajima, M. (2015). *R2jags: Using R to Run 'JAGS'*. R package version 0.5-7. 34

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, **10**, 277–303. 6, 10, 11, 12

Thompson, S. G. and Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, **21**, 1559–1573. 17

Thompson, S. G. and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine*, **18**, 2693–2708. 14, 16, 19

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **36**, 1–48. 7, 16, 33

West, D. B.*et al.* (2001). *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River. 25

White, I. R., Barrett, J. K., Jackson, D., and Higgins, J. (2012). Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods*, **3**, 111–125. 41, 46

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in medicine*, **10**, 1665–1677. 4, 7